Photonic Devices

Jia-Ming Liu

Professor of Electrical Engineering
University of California, Los Angeles

CAMBRIDGE
UNIVERSITY PRESS

Photonic Devices

Photonic devices lie at the heart of the communications revolution, and have become a large and important part of the electronic engineering field, so much so that many colleges now treat this as a subject in its own right. With this in mind, the author has put together a unique textbook covering every major photonic device, and striking a careful balance between theoretical and practical concepts. The book assumes a basic knowledge of optics, semiconductors, and electromagnetic waves; many of the key background concepts are reviewed in the first chapter. Devices covered include optical fibers, couplers, electro-optic devices, magneto-optic devices, acousto-optic devices, nonlinear optical devices, optical amplifiers, lasers, light-emitting diodes, and photodetectors. Problems are included at the end of each chapter and a solutions set is available. The book is ideal for senior undergraduate and graduate courses, but being device-driven it is also an excellent reference for engineers.

Jia-Ming Liu is Professor of Electrical Engineering at the University of California, Los Angeles. He received his Ph.D. degree in applied physics from Harvard University in 1982. His research interests are in the areas of nonlinear optics, ultrafast optics, photonic devices, optical wave propagation, nonlinear laser dynamics, and chaotic communications. Dr. Liu has written about 200 scientific publications and holds eight US patents. He is a fellow of the Optical Society of America and the American Physical Society.

1 General background

Photonics is an engineering discipline concerning the control of light, or photons, for useful applications, much as electronics has to do with electrons. Light is electromagnetic radiation of frequencies in the range from 1 THz to 10 PHz, corresponding to wavelengths between ~300 μm and ~30 nm in free space. This optical spectral range is generally divided into infrared, visible, and ultraviolet regions, as indicated in Table 1.1. The spectral range of concern in photonics is usually in a wavelength range between 10 μm and 100 nm. The primary interest in the applications of photonic devices is in an even narrower range of visible and near infrared wavelengths. As we shall see later, this spectral range of application is largely determined by the properties of materials used for photonic devices.

The wave nature of light is very important in the function of photonic devices. In particular, the propagation of light in a photonic device is completely characterized by its wave nature. However, in the spectral range of interest for practical photonic devices, the quantum energies of photons are in a range where the quantum nature of light is also important. For example, photons of visible light have energies between 1.7 and 3.1 eV, which are in the range of the bandgaps of most semiconductors. Photon energy is an important factor that determines the behavior of an optical wave in a semiconductor photonic device. The uniqueness of photonic devices is that both wave and quantum characteristics of light have to be considered for the function and applications of these devices. Generally speaking, the photon nature of light is important in the operation of photonic devices for generation, amplification, frequency conversion, or detection of light, while the wave nature is important in the operation of all photonic devices but is particularly so for devices used in transmission, modulation, or switching of light. In this chapter, we review some relevant wave and quantum properties of light as a general background for later chapters.

1.1 Optical fields and Maxwell's equations

When dealing with photonic devices, we consider in most situations optical fields in media of various electromagnetic properties. The electromagnetic field in a medium is

Table 1.1 *Electromagnetic spectrum*

Wave region	Frequency	Wavelength	Devices
Radio	kHz–MHz–GHz	km–m–cm	Electronic devices
Microwave	1 GHz–1 THz	300 mm–300 μm	Microwave devices
Optical			
Infrared	1 THz–430 THz	300 μm–700 nm	Photonic devices
Visible	430 THz–750 THz	700 nm–400 nm	
Ultraviolet	750 THz–10 PHz	400 nm–30 nm	
X-ray	10 PHz–10 EHz	30 nm–300 pm	
Gamma ray	10 EHz and above	300 pm and shorter	

generally characterized by the following four field quantities:

electric field $\boldsymbol{E}(\mathbf{r}, t)$ V m^{-1},
electric displacement $\boldsymbol{D}(\mathbf{r}, t)$ C m^{-2},
magnetic field $\boldsymbol{H}(\mathbf{r}, t)$ A m^{-1},
magnetic induction $\boldsymbol{B}(\mathbf{r}, t)$ T or Wb m^{-2}.

Note that $\boldsymbol{E}$ and $\boldsymbol{B}$ are fundamental microscopic fields, while $\boldsymbol{D}$ and $\boldsymbol{H}$ are macroscopic fields that include the response of the medium. The units given above and below for the field quantities are SI units consistent with the SI system used in this book for Maxwell's equations. Experimentally measured magnetic field quantities are sometimes given in Gaussian units, which are gauss for the $\boldsymbol{B}$ field and oersted (Oe) for the $\boldsymbol{H}$ field. The conversion relations between SI and Gaussian units are 1 T = 1 Wb m^{-2} = 10^4 gauss for $\boldsymbol{B}$ and 1 A m^{-1} = $4\pi \times 10^{-3}$ Oe for $\boldsymbol{H}$.

The response of a medium to an electromagnetic field generates the *polarization* and the *magnetization*:

polarization (electric polarization) $\boldsymbol{P}(\mathbf{r}, t)$ C m^{-2},
magnetization (magnetic polarization) $\boldsymbol{M}(\mathbf{r}, t)$ A m^{-1}.

They are connected to the field quantities through the following relations:

$$\boldsymbol{D}(\mathbf{r}, t) = \epsilon_0 \boldsymbol{E}(\mathbf{r}, t) + \boldsymbol{P}(\mathbf{r}, t) \tag{1.1}$$

and

$$\boldsymbol{B}(\mathbf{r}, t) = \mu_0 \boldsymbol{H}(\mathbf{r}, t) + \mu_0 \boldsymbol{M}(\mathbf{r}, t), \tag{1.2}$$

where

$$\epsilon_0 \approx \frac{1}{36\pi} \times 10^{-9} \text{ F m}^{-1} \text{ or A s V}^{-1} \text{ m}^{-1} \tag{1.3}$$

is the *electric permittivity* of free space and

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{H\ m^{-1}}\ \text{or}\ \mathrm{V\ s\ A^{-1}\ m^{-1}} \tag{1.4}$$

is the *magnetic permeability* of free space. In addition, independent charge or current sources may exist:

charge density $\rho(\mathbf{r}, t)$ $\mathrm{C\ m^{-3}}$,
current density $\boldsymbol{J}(\mathbf{r}, t)$ $\mathrm{A\ m^{-2}}$.

In a medium, the behavior of a time-varying electromagnetic field is governed by the following space- and time-dependent macroscopic *Maxwell's equations*:

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad \text{Faraday's law,} \tag{1.5}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} \qquad \text{Ampere's law,} \tag{1.6}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = \rho \qquad \text{Coulomb's law,} \tag{1.7}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \qquad \text{absence of magnetic monopoles.} \tag{1.8}$$

The current and charge densities are constrained by the following *continuity equation*:

$$\boldsymbol{\nabla} \cdot \boldsymbol{J} + \frac{\partial \rho}{\partial t} = 0 \qquad \text{conservation of charge.} \tag{1.9}$$

In a medium free of sources, $\boldsymbol{J} = 0$ and $\rho = 0$. Then, Maxwell's equations are simply

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{1.10}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t}, \tag{1.11}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = 0, \tag{1.12}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0. \tag{1.13}$$

These are the equations normally used for optical fields because optical fields are usually not generated directly by free currents or free charges.

Transformation properties

Maxwell's equations and the continuity equation are the basic physical laws that govern the behavior of electromagnetic fields. They are invariant under the transformation of *space inversion*, in which the spatial vector $\mathbf{r}$ is changed to $\mathbf{r}' = -\mathbf{r}$, or $(x, y, z) \rightarrow (-x, -y, -z)$, and the transformation of *time reversal*, in which the time variable t is changed to $t' = -t$, or $t \rightarrow -t$. This means that the form of these equations is not changed when we perform the space-inversion transformation or the time-reversal transformation, or both together.

Different quantities in Maxwell's equations have different transformation properties. An understanding of these properties is important and leads to a fundamental appreciation of the difference between the characteristics of the electric and magnetic fields, which is the origin of the difference between the electric and magnetic symmetry properties of materials. It also helps in understanding many basic characteristics of the electro-optic, magneto-optic, and nonlinear optical properties of materials to be addressed in later chapters.

The electric field vectors, $\boldsymbol{E}$ and $\boldsymbol{D}$, have the same transformation properties as those of $\boldsymbol{P}$, while the transformation properties of the magnetic field vectors, $\boldsymbol{H}$ and $\boldsymbol{B}$, are the same as those of $\boldsymbol{M}$. The origin of the electric properties of a material is the charge-density distribution, $\rho(\mathbf{r}, t)$, at the atomic level in the material, whereas that of the magnetic properties stems from the current-density distribution, $\boldsymbol{J}(\mathbf{r}, t)$. The transformation properties of the scalar quantity ρ are such that *the sign of ρ remains unchanged under the transformation of either space inversion or time reversal*. In contrast, $\boldsymbol{J}$ is a *polar vector* because it is charge density times velocity, $\rho\mathbf{v}$, where velocity, $\mathbf{v}$, is a polar vector. Thus, *the vector $\boldsymbol{J}$ changes sign under the transformation of either space inversion or time reversal*. It changes sign under space inversion because a polar vector changes sign under space inversion, and it changes sign under time reversal because $\mathbf{v}$ is the first time derivative of $\mathbf{r}$. The electric polarization $\boldsymbol{P}$ is a polar vector because it is the volume average of the electric dipole moment density defined by $\rho(\mathbf{r}, t)\mathbf{r}$, and the product of a scalar quantity ρ and a polar vector $\mathbf{r}$ is a polar vector. In contrast, magnetization $\boldsymbol{M}$ is an *axial vector* because it is the volume average of the magnetic dipole moment density defined by $\mathbf{r} \times \boldsymbol{J}(\mathbf{r}, t)$, and the cross product of two polar vectors, $\mathbf{r}$ and $\boldsymbol{J}$, is an axial vector. Therefore, we find the following transformation properties.

1. **Electric fields.** The electric field vectors, $\boldsymbol{P}$, $\boldsymbol{E}$, and $\boldsymbol{D}$, change sign under space inversion but not under time reversal.
2. **Magnetic fields.** The magnetic field vectors, $\boldsymbol{M}$, $\boldsymbol{H}$, and $\boldsymbol{B}$ change sign under time reversal but not under space inversion.

With these transformation properties understood, the invariance of Maxwell's equations and the continuity equation under the transformation of space inversion or time reversal or both can be easily verified.

Response of medium

Polarization and magnetization in a medium are generated, respectively, by the response of the medium to the electric and magnetic fields. Therefore, $\boldsymbol{P}(\mathbf{r}, t)$ depends on $\boldsymbol{E}(\mathbf{r}, t)$, while $\boldsymbol{M}(\mathbf{r}, t)$ depends on $\boldsymbol{B}(\mathbf{r}, t)$. *At optical frequencies, the magnetization vanishes*, $\boldsymbol{M} = 0$. Consequently, for optical fields, the following relation is always true:

$$\boldsymbol{B}(\mathbf{r}, t) = \mu_0 \boldsymbol{H}(\mathbf{r}, t). \tag{1.14}$$

This is not true at low frequencies, however. It is possible to change the properties of a medium through a magnetization induced by a DC or low-frequency magnetic field, leading to the functioning of magneto-optic devices. It should be noted that even for magneto-optic devices, magnetization is induced by a DC or low-frequency magnetic field that is separate from the optical fields. No magnetization is induced by the magnetic components of the optical fields.

Except for magneto-optic devices, most photonic devices are made of dielectric materials that have zero magnetization at all frequencies. The optical properties of such materials are completely determined by the relation between $\boldsymbol{P}(\mathbf{r}, t)$ and $\boldsymbol{E}(\mathbf{r}, t)$. This relation is generally characterized by an *electric susceptibility tensor*, $\boldsymbol{\chi}$, through the following definition for electric polarization:

$$\boldsymbol{P}(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\chi}(\mathbf{r} - \mathbf{r}', t - t') \cdot \boldsymbol{E}(\mathbf{r}', t'). \tag{1.15}$$

From (1.1), then

$$\begin{aligned} \boldsymbol{D}(\mathbf{r}, t) &= \epsilon_0 \boldsymbol{E}(\mathbf{r}, t) + \epsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\chi}(\mathbf{r} - \mathbf{r}', t - t') \cdot \boldsymbol{E}(\mathbf{r}', t') \\ &= \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\epsilon}(\mathbf{r} - \mathbf{r}', t - t') \cdot \boldsymbol{E}(\mathbf{r}', t'), \end{aligned} \tag{1.16}$$

where $\boldsymbol{\epsilon}$ is the *electric permittivity tensor* of the medium.

Because $\boldsymbol{\chi}$ and, equivalently, $\boldsymbol{\epsilon}$ represent the response of a medium to the optical field and thus completely characterize the macroscopic electromagnetic properties of the medium, (1.15) and (1.16) can be regarded as the definitions of $\boldsymbol{P}(\mathbf{r}, t)$ and $\boldsymbol{D}(\mathbf{r}, t)$, respectively. A few remarks can be made:

1. Both $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ are generally tensors because the vectors $\boldsymbol{P}$ and $\boldsymbol{D}$ are, in general, not parallel to vector $\boldsymbol{E}$ due to material *anisotropy*. In the case of an *isotropic* medium, both $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ can be reduced to scalars χ and ϵ, respectively.
2. The relations in (1.15) and (1.16) are in the form of convolution integrals. The convolution in time accounts for the fact that the response of a medium to excitation of an electric field is generally not *instantaneous* or *local* in time and will not vanish for some time after the excitation is over. Because time is unidirectional, *causality* exists in physical processes. An earlier excitation can have an effect on the property of a medium at a later time, but not a later excitation on the property of the medium at an earlier time. Therefore, the upper limit in the time integral is t, not infinity. In contrast, the convolution in space accounts for the *spatial nonlocality* of the material response. Excitation of a medium at a location $\mathbf{r}'$ can result in a change

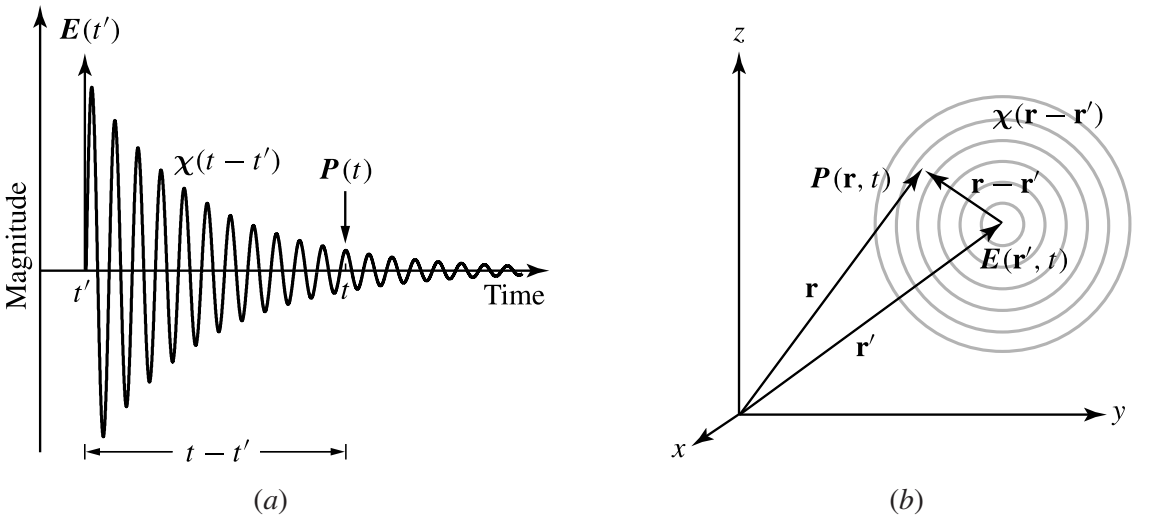


Figure 1.1 Nonlocal responses in (*a*) time and (*b*) space.

in the property of the medium at another location $\mathbf{r}$. For example, the property of a semiconductor at one location can be changed by electric or optical excitation at another location through carrier diffusion. Because space is not unidirectional, there is no spatial causality, in general, and spatial convolution is integrated over the entire space. Figure 1.1 shows the temporal and spatial nonlocality of responses to electromagnetic excitations. The temporal nonlocality of the optical response of a medium results in *frequency dispersion* of its optical property, while the spatial nonlocality results in *momentum dispersion*.

3. In addition to the dependence on space and time through the convolution relation with the optical field, χ and ϵ can also be functions of space or time independent of the optical field because of spatial or temporal *inhomogeneities* in the medium. Spatial inhomogeneity exists in all optical structures, such as optical waveguides, where the index of refraction is a function of space. Temporal inhomogeneity exists when the optical property of a medium varies with time, for example, because of modulation by a low-frequency electric field or by an acoustic wave.
4. In a linear medium, χ and ϵ do not depend on the optical field $\boldsymbol{E}$. In a nonlinear optical material, χ and ϵ are themselves also functions of $\boldsymbol{E}$.

Boundary conditions

At the interface of two media of different optical properties as shown in Fig. 1.2, the optical field components must satisfy certain boundary conditions. These boundary conditions can be derived from Maxwell's equations given in (1.10)–(1.13). From (1.10) and (1.11), the tangential components of the fields at the boundary satisfy

$$\hat{n} \times \boldsymbol{E}_1 = \hat{n} \times \boldsymbol{E}_2 \tag{1.17}$$

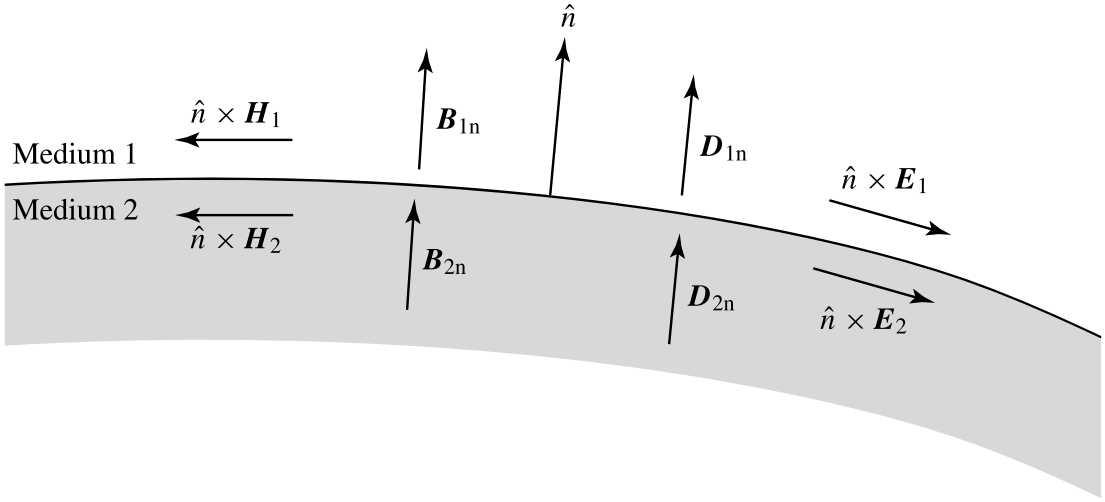


Figure 1.2 Boundary between two media of different optical properties.

and

$$\hat{n} \times \boldsymbol{H}_1 = \hat{n} \times \boldsymbol{H}_2, \tag{1.18}$$

where $\hat{n}$ is the unit vector normal to the interface as shown in Fig. 1.2. From (1.12) and (1.13), we have

$$\hat{n} \cdot \boldsymbol{D}_1 = \hat{n} \cdot \boldsymbol{D}_2 \tag{1.19}$$

and

$$\hat{n} \cdot \boldsymbol{B}_1 = \hat{n} \cdot \boldsymbol{B}_2 \tag{1.20}$$

for the normal components of the fields.

The tangential components of $\boldsymbol{E}$ and $\boldsymbol{H}$ must be continuous across an interface, while the normal components of $\boldsymbol{D}$ and $\boldsymbol{B}$ are continuous. Because $\boldsymbol{B} = \mu_0 \boldsymbol{H}$ for optical fields, as discussed above, (1.18) and (1.20) also imply that the tangential component of $\boldsymbol{B}$ and the normal component of $\boldsymbol{H}$ are also continuous. Consequently, *all of the magnetic field components in an optical field are continuous across a boundary. Possible discontinuities in an optical field exist only in the normal component of* $\boldsymbol{E}$ *or the tangential component of* $\boldsymbol{D}$.

Optical power and energy

By multiplying $\boldsymbol{E}$ by (1.6) and multiplying $\boldsymbol{H}$ by (1.5), we obtain

$$\boldsymbol{E} \cdot (\boldsymbol{\nabla} \times \boldsymbol{H}) = \boldsymbol{E} \cdot \boldsymbol{J} + \boldsymbol{E} \cdot \frac{\partial \boldsymbol{D}}{\partial t}, \tag{1.21}$$

$$\boldsymbol{H} \cdot (\boldsymbol{\nabla} \times \boldsymbol{E}) = -\boldsymbol{H} \cdot \frac{\partial \boldsymbol{B}}{\partial t}. \tag{1.22}$$

Using the vector identity

$$\mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) = \nabla \cdot (\mathbf{A} \times \mathbf{B}), \tag{1.23}$$

we can combine (1.21) and (1.22) to have

$$-\nabla \cdot (\boldsymbol{E} \times \boldsymbol{H}) = \boldsymbol{E} \cdot \boldsymbol{J} + \boldsymbol{E} \cdot \frac{\partial \boldsymbol{D}}{\partial t} + \boldsymbol{H} \cdot \frac{\partial \boldsymbol{B}}{\partial t}. \tag{1.24}$$

Using (1.1) and (1.2) and rearranging (1.24), we obtain

$$\boldsymbol{E} \cdot \boldsymbol{J} = -\nabla \cdot (\boldsymbol{E} \times \boldsymbol{H}) - \frac{\partial}{\partial t}\left(\frac{\epsilon_0}{2}|\boldsymbol{E}|^2 + \frac{\mu_0}{2}|\boldsymbol{H}|^2\right) - \left(\boldsymbol{E} \cdot \frac{\partial \boldsymbol{P}}{\partial t} + \mu_0 \boldsymbol{H} \cdot \frac{\partial \boldsymbol{M}}{\partial t}\right). \tag{1.25}$$

Recall that power in an electric circuit is given by voltage times current and has the unit of W = V A (watts = volts × amperes). In an electromagnetic field, we find similarly that $\boldsymbol{E} \cdot \boldsymbol{J}$ is the power density that has the unit of V A m^{-3} or W m^{-3}. Therefore, the total power dissipated by an electromagnetic field in a volume $\mathcal{V}$ is just

$$\int_{\mathcal{V}} \boldsymbol{E} \cdot \boldsymbol{J}\, \mathrm{d}\mathcal{V}. \tag{1.26}$$

Expressing (1.25) in an integral form, we have

$$\begin{aligned}\int_{\mathcal{V}} \boldsymbol{E} \cdot \boldsymbol{J}\, \mathrm{d}\mathcal{V} = &- \oint_{\mathcal{A}} \boldsymbol{E} \times \boldsymbol{H} \cdot \hat{n}\, \mathrm{d}\mathcal{A} - \frac{\partial}{\partial t}\int_{\mathcal{V}} \left(\frac{\epsilon_0}{2}|\boldsymbol{E}|^2 + \frac{\mu_0}{2}|\boldsymbol{H}|^2\right) \mathrm{d}\mathcal{V} \\ &- \int_{\mathcal{V}} \left(\boldsymbol{E} \cdot \frac{\partial \boldsymbol{P}}{\partial t} + \mu_0 \boldsymbol{H} \cdot \frac{\partial \boldsymbol{M}}{\partial t}\right) \mathrm{d}\mathcal{V},\end{aligned} \tag{1.27}$$

where the first term on the right-hand side is a surface integral over the closed surface $\mathcal{A}$ of volume $\mathcal{V}$ and $\hat{n}$ is the outward-pointing unit normal vector of the surface, as shown in Fig. 1.3.

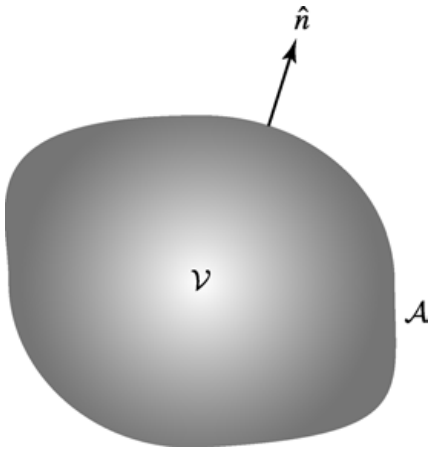


Figure 1.3 Boundary surface enclosing a volume element and the unit surface normal vector.

Clearly, each term in (1.27) has the unit of power. Each has an important physical meaning. The vector quantity

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \tag{1.28}$$

is called the *Poynting vector* of the electromagnetic field. It represents the *instantaneous magnitude and direction of the power flow* of the field. The scalar quantity

$$u_0 = \frac{\epsilon_0}{2}|\boldsymbol{E}|^2 + \frac{\mu_0}{2}|\boldsymbol{H}|^2 \tag{1.29}$$

has the unit of energy per unit volume and is the *energy density stored in the propagating field*. It consists of two components, thus accounting for energies stored in both electric and magnetic fields at any instant of time. The last term in (1.27) also has two components associated with electric and magnetic fields, respectively. The quantity

$$W_{\mathrm{p}} = \boldsymbol{E} \cdot \frac{\partial \boldsymbol{P}}{\partial t} \tag{1.30}$$

is the *power density expended by the electromagnetic field on the polarization*. It is the rate of energy transfer from the electromagnetic field to the medium by inducing electric polarization in the medium. Similarly, the quantity

$$W_{\mathrm{m}} = \mu_0 \boldsymbol{H} \cdot \frac{\partial \boldsymbol{M}}{\partial t} \tag{1.31}$$

is the *power density expended by the electromagnetic field on the magnetization.* With these physical meanings attached to these terms, it can be seen that (1.27) simply states the law of conservation of energy in any arbitrary volume element $\mathcal{V}$ in the medium. The total energy in the medium equals that in the propagating field plus that in the electric and magnetic polarizations.

In the special case of a linear, nondispersive medium where $\boldsymbol{\epsilon}(\mathbf{r} - \mathbf{r}', t - t') = \boldsymbol{\epsilon}\delta(\mathbf{r} - \mathbf{r}')\delta(t - t')$, (1.16) simply reduces to $\boldsymbol{D}(\mathbf{r}, t) = \boldsymbol{\epsilon} \cdot \boldsymbol{E}(\mathbf{r}, t)$. Then, instead of (1.25), we have

$$\boldsymbol{E} \cdot \boldsymbol{J} = -\boldsymbol{\nabla} \cdot \boldsymbol{S} - \frac{\partial}{\partial t}\left(\frac{1}{2}\boldsymbol{E} \cdot \boldsymbol{D} + \frac{1}{2}\boldsymbol{H} \cdot \boldsymbol{B}\right) \tag{1.32}$$

from (1.24). In this situation, the total energy density stored in the medium, including that in the propagating field and that in the polarizations, is simply

$$u = \frac{1}{2}\boldsymbol{E} \cdot \boldsymbol{D} + \frac{1}{2}\boldsymbol{H} \cdot \boldsymbol{B}. \tag{1.33}$$

For an optical field, $\boldsymbol{J} = 0$ and $\boldsymbol{M} = 0$, as is discussed above. Then, (1.27) becomes

$$-\oint_{\mathcal{A}} \boldsymbol{S} \cdot \hat{n}\mathrm{d}\mathcal{A} = \frac{\partial}{\partial t}\int_{\mathcal{V}} u_0 \mathrm{d}\mathcal{V} + \int_{\mathcal{V}} W_{\mathrm{p}} \mathrm{d}\mathcal{V}, \tag{1.34}$$

which states that the total optical power flowing into volume $\mathcal{V}$ through its boundary surface $\mathcal{A}$ is equal to the rate of increase with time of the energy stored in the propagating fields in $\mathcal{V}$ plus the power transferred to the polarization of the medium in this volume. In a linear, nondispersive medium, we have

$$-\oint_{\mathcal{A}} \boldsymbol{S}\cdot\hat{n}\mathrm{d}\mathcal{A} = \frac{\partial}{\partial t}\int_{\mathcal{V}} u\mathrm{d}\mathcal{V}. \tag{1.35}$$

Wave equation

By applying $\boldsymbol{\nabla}\times$ to (1.10) and using (1.14) and (1.11), we have

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\boldsymbol{E} + \mu_0\frac{\partial^2\boldsymbol{D}}{\partial t^2} = 0. \tag{1.36}$$

Using (1.1), (1.36) can be expressed as

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\boldsymbol{E} + \frac{1}{c^2}\frac{\partial^2\boldsymbol{E}}{\partial t^2} = -\mu_0\frac{\partial^2\boldsymbol{P}}{\partial t^2}, \tag{1.37}$$

where

$$c = \frac{1}{\sqrt{\mu_0\epsilon_0}} \approx 3\times 10^8 \text{ m s}^{-1} \tag{1.38}$$

is the *speed of light* in free space. The *wave equation* in (1.37) describes the space-and-time evolution of the electric field of the optical wave. Its right-hand side can be regarded as the driving source for the optical wave. The polarization in a medium drives the evolution of an optical field. This wave equation can take on various forms depending on the characteristics of the medium, as will be seen on various occasions later. For now, we leave it in this general form.

1.2 Harmonic fields

Optical fields are harmonic fields that vary sinusoidally with time. The field vectors defined in the preceding section are all real quantities. For harmonic fields, it is always convenient to use *complex fields*. We define the space- and time-dependent complex electric field, $\mathbf{E}(\mathbf{r}, t)$, through its relation to the real electric field, $\boldsymbol{E}(\mathbf{r}, t)$:[1]

$$\boldsymbol{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) + \mathbf{E}^*(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) + \text{c.c.}, \tag{1.39}$$

[1] In some literature, the complex field is defined through a relation with the real field as $\boldsymbol{E}(\mathbf{r}, t) = 1/2(\mathbf{E}(\mathbf{r}, t) + \mathbf{E}^*(\mathbf{r}, t))$, which differs from the relation in (1.39) by the factor $1/2$. The magnitude of the complex field defined through this alternative relation is twice that of the complex field defined through (1.39). As a result, expressions for many quantities may be different under the two different definitions. An example is that of the optical intensity given in (1.98). We have chosen to define the complex field through the relation in (1.39) without the factor $1/2$ primarily because this definition is more convenient and less confusing in expressing the nonlinear polarizations discussed in Chapter 9.

where c.c. means the complex conjugate. In our convention, $\mathbf{E}(\mathbf{r}, t)$ contains the complex field components that vary with time as $\exp(-\mathrm{i}\omega t)$ with positive values of ω, while $\mathbf{E}^*(\mathbf{r}, t)$ contains those varying with time as $\exp(\mathrm{i}\omega t)$ with positive ω, or $\exp(-\mathrm{i}\omega t)$ with negative ω. The complex fields of other field quantities are similarly defined (see Appendix A).

With this definition for the complex fields, all of the linear field equations retain their forms. In particular, Maxwell's equations for the complex optical fields are

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{1.40}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \tag{1.41}$$

$$\nabla \cdot \mathbf{D} = 0, \tag{1.42}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{1.43}$$

The wave equation in terms of the complex electric field is

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}, \tag{1.44}$$

while

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\chi}(\mathbf{r} - \mathbf{r}', t - t') \cdot \mathbf{E}(\mathbf{r}', t') \tag{1.45}$$

and

$$\begin{aligned} \mathbf{D}(\mathbf{r}, t) &= \epsilon_0 \mathbf{E}(\mathbf{r}, t) + \epsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\chi}(\mathbf{r} - \mathbf{r}', t - t') \cdot \mathbf{E}(\mathbf{r}', t') \\ &= \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}' \int_{-\infty}^{t} \mathrm{d}t' \boldsymbol{\epsilon}(\mathbf{r} - \mathbf{r}', t - t') \cdot \mathbf{E}(\mathbf{r}', t'). \end{aligned} \tag{1.46}$$

It is important to note that while $\mathbf{P}$, $\mathbf{D}$, and $\mathbf{E}$ are complex, $\boldsymbol{\chi}(\mathbf{r} - \mathbf{r}', t - t')$ and $\boldsymbol{\epsilon}(\mathbf{r} - \mathbf{r}', t - t')$ in (1.45) and (1.46) are always real and are the same as those in (1.15) and (1.16).

For a harmonic optical field of wavevector $\mathbf{k}$ and angular frequency ω, its complex electric field can be further written as

$$\mathbf{E}(\mathbf{r}, t) = \boldsymbol{\mathcal{E}}(\mathbf{r}, t) \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{r} - \mathrm{i}\omega t), \tag{1.47}$$

where $\boldsymbol{\mathcal{E}}(\mathbf{r}, t)$ is the space- and time-varying field envelope, such as that for a modulated field, a guided field, or an optical pulse. Other complex field quantities, such as $\mathbf{H}(\mathbf{r}, t)$, can be similarly expressed. The phase factor in (1.47) indicates the direction

of wave propagation:

$$\begin{aligned} &\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\mathrm{i}\omega t, \quad \text{forward propagating in } \mathbf{k} \text{ direction;} \\ &-\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\mathrm{i}\omega t, \quad \text{backward propagating in } -\mathbf{k} \text{ direction.} \end{aligned}$$

The light *intensity*, or *irradiance*, is the *power density* of the harmonic optical field. It can be calculated by time averaging of the Poynting vector over one wave cycle:

$$\overline{S} = \frac{1}{T}\int_0^T \boldsymbol{E}\times\boldsymbol{H}\mathrm{d}t = 2\,\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*), \tag{1.48}$$

where Re(·) means taking the real part. We can define a *complex Poynting vector*:

$$\mathbf{S} = \mathbf{E}\times\mathbf{H}^* \tag{1.49}$$

so that

$$\overline{S} = \mathbf{S}+\mathbf{S}^*, \tag{1.50}$$

which has the same form as the relation between the real and complex fields defined in (1.39) except that the real Poynting vector in this relation is time averaged. The light intensity, I, is simply the magnitude of the *real* time-averaged Poynting vector:

$$I = |\overline{S}| = |\mathbf{S}+\mathbf{S}^*|, \tag{1.51}$$

where I is in watts per square meter.

For harmonic optical fields, it is often useful to consider the complex fields in the momentum space and frequency domain defined by the following Fourier-transform relations:

$$\mathbf{E}(\mathbf{k},\omega) = \int_{-\infty}^{\infty}\mathrm{d}\mathbf{r}\int_{-\infty}^{\infty}\mathrm{d}t\mathbf{E}(\mathbf{r},t)\exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{r}+\mathrm{i}\omega t), \quad \text{for } \omega>0, \tag{1.52}$$

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{k}\int_0^{\infty}\mathrm{d}\omega\mathbf{E}(\mathbf{k},\omega)\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\mathrm{i}\omega t). \tag{1.53}$$

Note that $\mathbf{E}(\mathbf{k},\omega)$ in (1.52) is defined for $\omega>0$ only, and the integral for the time dependence of $\mathbf{E}(\mathbf{r},t)$ in (1.53) extends only over positive values of ω. This is in accordance with the convention we used to define the complex field $\mathbf{E}(\mathbf{r},t)$ in (1.39). All other space- and time-dependent quantities, including other field vectors and the permittivity and susceptibility tensors, are transformed in a similar manner. Through the Fourier transform, the convolution integrals in real space and time become simple products in the momentum space and frequency domain. Consequently,

we have

$$\mathbf{P}(\mathbf{k}, \omega) = \epsilon_0 \boldsymbol{\chi}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{k}, \omega) \tag{1.54}$$

and

$$\mathbf{D}(\mathbf{k}, \omega) = \boldsymbol{\epsilon}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{k}, \omega). \tag{1.55}$$

1.3 Linear optical susceptibility

As mentioned above, the susceptibility tensor $\boldsymbol{\chi}(\mathbf{r}, t)$ and the permittivity tensor $\boldsymbol{\epsilon}(\mathbf{r}, t)$ of space and time are always real quantities although all field quantities, including both $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{k}, \omega)$, can be defined in a complex form. This is true even in the presence of an optical loss or gain in the medium. However, the susceptibility and permittivity tensors in the momentum space and frequency domain, $\boldsymbol{\chi}(\mathbf{k}, \omega)$ and $\boldsymbol{\epsilon}(\mathbf{k}, \omega)$, can be complex. If an eigenvalue, χ_i, of $\boldsymbol{\chi}$ is complex, the corresponding eigenvalue, ϵ_i, of $\boldsymbol{\epsilon}$ is also complex, and their imaginary parts have the same sign because $\boldsymbol{\epsilon} = \epsilon_0(1 + \boldsymbol{\chi})$. The signs of such imaginary parts of eigenvalues tell whether the medium has an optical gain or loss. In our convention, we write, for example, $\chi_i = \chi_i' + \mathrm{i}\chi_i''$ in the frequency domain. Then, $\chi_i''(\omega) > 0$ *corresponds to an optical loss or absorption*, while $\chi_i''(\omega) < 0$ *represents an optical gain or amplification*.

The fact that $\boldsymbol{\chi}(\mathbf{r}, t)$ and $\boldsymbol{\epsilon}(\mathbf{r}, t)$ are real quantities leads to the following symmetry relations for the tensor elements of $\boldsymbol{\chi}(\mathbf{k}, \omega)$ and $\boldsymbol{\epsilon}(\mathbf{k}, \omega)$:

$$\chi_{ij}^*(\mathbf{k}, \omega) = \chi_{ij}(-\mathbf{k}, -\omega) \tag{1.56}$$

and

$$\epsilon_{ij}^*(\mathbf{k}, \omega) = \epsilon_{ij}(-\mathbf{k}, -\omega), \tag{1.57}$$

which are called the *reality condition*. The reality condition implies that $\chi_{ij}'(\mathbf{k}, \omega) = \chi_{ij}'(-\mathbf{k}, -\omega)$ and $\chi_{ij}''(\mathbf{k}, \omega) = -\chi_{ij}''(-\mathbf{k}, -\omega)$. Similar relations also apply for the real and imaginary parts of ϵ_{ij}. Therefore, the real parts of χ_{ij} and ϵ_{ij} are even functions of $\mathbf{k}$ and ω, whereas the imaginary parts are odd functions of $\mathbf{k}$ and ω. Any constant contribution, independent of $\mathbf{k}$ and ω, in χ_{ij} and ϵ_{ij} is an even function of $\mathbf{k}$ and ω; hence it can appear only in the real parts. As a result, the imaginary parts, if they exist, are always functions of either $\mathbf{k}$ or ω, or both. The loss, or gain, in a medium is associated with the imaginary parts of the eigenvalues of $\boldsymbol{\chi}(\omega)$; consequently, it is inherently dispersive. Any other effects that can be described by the imaginary parts of the eigenvalues of $\boldsymbol{\chi}(\mathbf{k}, \omega)$ are also dispersive in either momentum or frequency, or both.

The momentum and frequency dependencies of an electric susceptibility, $\boldsymbol{\chi}(\mathbf{k}, \omega)$, are due to the spatial and temporal nonlocality properties of the underlying physical

mechanisms that contribute to χ. As discussed in the preceding section, spatial nonlocality causes spatially convoluted effects and results in momentum dependence of the susceptibility, and temporal nonlocality causes temporal convolution and results in frequency dispersion of the medium.

In addition to nonlocality, it is also important to consider inhomogeneity, in both space and time. *In a linear medium, changes in the wavevector of an optical wave, or coupling between waves of different wavevectors, can occur only if the optical property of the medium in which the wave propagates is spatially inhomogeneous such that* $\chi(\mathbf{k}, \omega)$ *is spatially dependent*. Likewise, *changes in the frequency of an optical wave, or coupling between waves of different frequencies, are possible in a linear medium only if the optical property of the medium is time varying such that* $\chi(\mathbf{k}, \omega)$ *varies with time.* Changes in the wavevector of an optical wave can take the form of changes in the wave propagation direction, as in reflection and diffraction, or in the optical wavelength, as in the case when a wave propagates from one part of the medium to another of different refractive index. Changes in the frequency of an optical wave result in the generation of other frequencies or the conversion of the optical wave to a completely different frequency. Consequently, for practical photonic devices, it is often necessary to consider both nonlocality and inhomogeneity in both space and time, thus writing $\chi(\mathbf{r}, t; \mathbf{k}, \omega)$ and, correspondingly, $\epsilon(\mathbf{r}, t; \mathbf{k}, \omega)$.

1.4 Polarization of light

Consider a monochromatic plane optical wave that has a complex field

$$\mathbf{E}(\mathbf{r}, t) = \boldsymbol{\mathcal{E}} \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{r} - \mathrm{i}\omega t) = \hat{e}\mathcal{E} \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{r} - \mathrm{i}\omega t), \tag{1.58}$$

where $\boldsymbol{\mathcal{E}}$ is a constant independent of $\mathbf{r}$ and t, and $\hat{e}$ is its unit vector. The *polarization* of the optical field is characterized by the unit vector $\hat{e}$. The wave is *linearly polarized*, also called *plane polarized*, if $\hat{e}$ can be expressed as a constant, real vector. Otherwise, the wave is *elliptically polarized* in general, and is *circularly polarized* in some special cases. For the convenience of discussion, we take the direction of wave propagation to be the z direction so that $\mathbf{k} = k\hat{z}$ and assume that both $\mathbf{E}$ and $\mathbf{H}$ lie in the xy plane.[2] Then, we have

$$\boldsymbol{\mathcal{E}} = \hat{x}\mathcal{E}_x + \hat{y}\mathcal{E}_y = \hat{x}|\mathcal{E}_x|\mathrm{e}^{\mathrm{i}\varphi_x} + \hat{y}|\mathcal{E}_y|\mathrm{e}^{\mathrm{i}\varphi_y}, \tag{1.59}$$

where $\mathcal{E}_x$ and $\mathcal{E}_y$ are space- and time-independent complex amplitudes, with phases φ_x and φ_y, respectively.

[2] This assumption is generally true if the medium is isotropic. It is not necessarily true if the medium is anisotropic. Propagation and polarization in isotropic and anisotropic media are discussed in the following two sections. However, the general concept discussed here does not depend on the validity of this assumption.

The polarization of the wave depends only on the phase difference and the magnitude ratio between the two field components $\mathcal{E}_x$ and $\mathcal{E}_y$. It can be completely characterized by the following two parameters:

$$\varphi = \varphi_y - \varphi_x, \qquad -\pi < \varphi \leq \pi, \tag{1.60}$$

and

$$\alpha = \tan^{-1} \frac{|\mathcal{E}_y|}{|\mathcal{E}_x|}, \qquad 0 \leq \alpha \leq \frac{\pi}{2}. \tag{1.61}$$

Because only the relative phase φ matters, we can set $\varphi_x = 0$ and take $\mathcal{E}$ to be real in the following discussions. Then $\boldsymbol{\mathcal{E}}$ from (1.59) can be written as

$$\boldsymbol{\mathcal{E}} = \mathcal{E}\hat{e}, \quad \text{with} \quad \hat{e} = \hat{x}\cos\alpha + \hat{y}e^{i\varphi}\sin\alpha. \tag{1.62}$$

Using (1.39), the space- and time-dependent real field is

$$\boldsymbol{E}(z,t) = 2\mathcal{E}\left[\hat{x}\cos\alpha\cos(kz - \omega t) + \hat{y}\sin\alpha\cos(kz - \omega t + \varphi)\right]. \tag{1.63}$$

At a fixed z location, say $z = 0$, we see that the electric field varies with time as

$$\boldsymbol{E}(t) = 2\mathcal{E}\left[\hat{x}\cos\alpha\cos\omega t + \hat{y}\sin\alpha\cos(\omega t - \varphi)\right]. \tag{1.64}$$

In general, $\mathcal{E}_x$ and $\mathcal{E}_y$ have different phases and different magnitudes. Therefore, the values of φ and α can be any combination. At a fixed point in space, both the direction and the magnitude of the field vector $\boldsymbol{E}$ in (1.64) can vary with time. Except when the values of φ and α fall into one of the special cases discussed below, the tip of this vector generally describes an ellipse, and the wave is said to be elliptically polarized. Note that we have assumed that the wave propagates in the positive z direction. When we view the ellipse by facing *against* this direction of wave propagation, we see that the tip of the field vector rotates *counterclockwise*, or *left handedly*, if $\varphi > 0$, and *clockwise*, or *right handedly*, if $\varphi < 0$. Figure 1.4 shows the ellipse traced by the tip of the rotating field vector at a fixed point in space. Also shown in the figure are the relevant parameters that characterize elliptic polarization.

In the description of the polarization characteristics of an optical wave, it is sometimes convenient to use, in place of α and φ, a set of two other parameters, θ and ε, which specify the *orientation* and *ellipticity* of the ellipse, respectively. The orientational parameter θ is the directional angle measured from the x axis to the major axis of the ellipse. Its range is taken to be $0 \leq \theta < \pi$ for convenience. Ellipticity ε is defined as

$$\varepsilon = \pm\tan^{-1}\frac{b}{a}, \qquad -\frac{\pi}{4} \leq \varepsilon \leq \frac{\pi}{4}, \tag{1.65}$$

where a and b are the major and minor semiaxes, respectively, of the ellipse. The plus

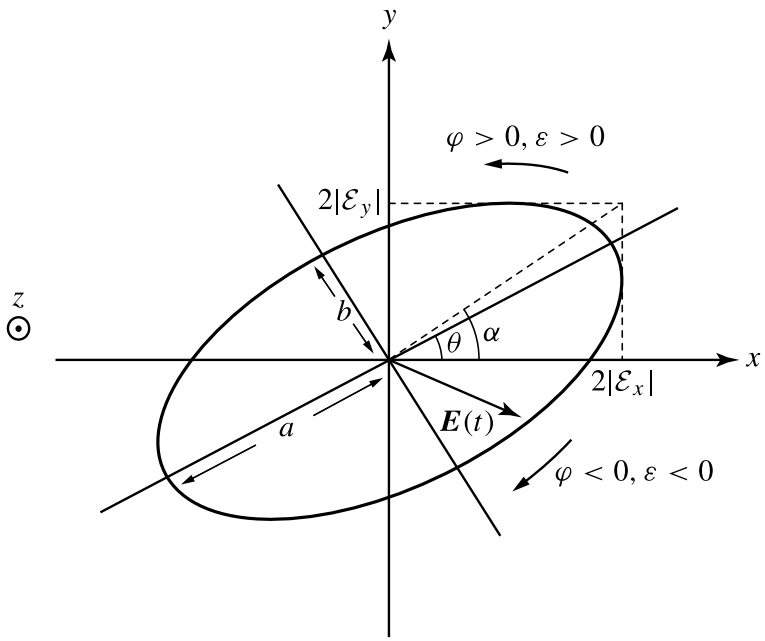


Figure 1.4 Ellipse described by the tip of the field of an elliptically polarized optical wave at a fixed point in space. Also shown are relevant parameters characterizing the state of polarization. The propagation direction is assumed to be the positive z direction, and the ellipse is viewed by facing against this direction.

sign for $\varepsilon > 0$ is taken to correspond to $\varphi > 0$ for left-handed polarization, whereas the minus sign for $\varepsilon < 0$ is taken to correspond to $\varphi < 0$ for right-handed polarization. The two sets of parameters (α, φ) and (θ, ε) have the following relations:

$$\tan 2\theta = \tan 2\alpha \cos\varphi, \tag{1.66}$$

$$\sin 2\varepsilon = \sin 2\alpha \sin\varphi. \tag{1.67}$$

Either set is sufficient to characterize the polarization state of an optical wave completely.

The following special cases are of particular interest.

1. **Linear polarization.** This happens when $\varphi = 0$ or π for any value of α. It is also characterized by $\varepsilon = 0$, and $\theta = \alpha$, if $\varphi = 0$, or $\theta = \pi - \alpha$, if $\varphi = \pi$. Clearly, the ratio $\mathcal{E}_x/\mathcal{E}_y$ is real in this case; therefore, linear polarization is described by a constant, real unit vector as

   $$\hat{e} = \hat{x}\cos\theta + \hat{y}\sin\theta. \tag{1.68}$$

   It follows that $\boldsymbol{E}(t)$ described by (1.64) reduces to

   $$\boldsymbol{E}(t) = 2\mathcal{E}\hat{e}\cos\omega t, \tag{1.69}$$

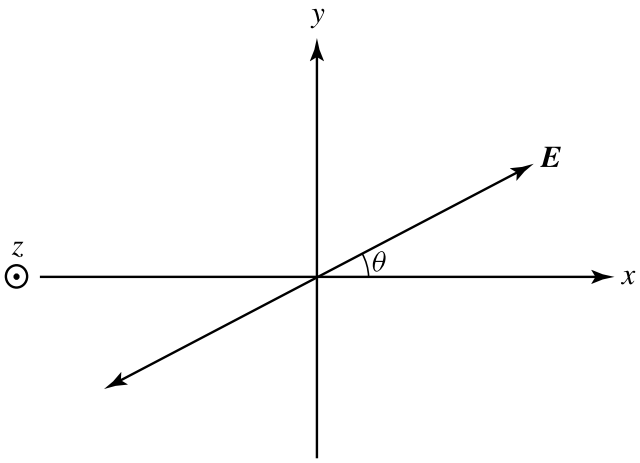


Figure 1.5 Field of a linearly polarized optical wave.

The tip of this vector traces a line in space at an angle θ with respect to the x axis, as shown in Fig. 1.5.

2. **Circular polarization.** This happens when $\varphi = \pi/2$ or $-\pi/2$, and $\alpha = \pi/4$. It is also characterized by $\varepsilon = \pi/4$ or $-\pi/4$, and $\theta = 0$. Because $\alpha = \pi/4$, we have $|\mathcal{E}_x| = |\mathcal{E}_y| = \mathcal{E}/\sqrt{2}$. There are two different circular polarization states:

   a. **Left-circular polarization.** For $\varphi = \pi/2$, also $\varepsilon = \pi/4$, the wave is *left-circularly polarized* if it propagates in the positive z direction. The complex field amplitude in (1.62) becomes

   $$\boldsymbol{\mathcal{E}} = \mathcal{E}\frac{\hat{x} + \mathrm{i}\hat{y}}{\sqrt{2}} = \mathcal{E}\hat{e}_+, \tag{1.70}$$

   and $\boldsymbol{E}(t)$ described by (1.64) reduces to

   $$\boldsymbol{E}(t) = \sqrt{2}\mathcal{E}(\hat{x}\cos\omega t + \hat{y}\sin\omega t). \tag{1.71}$$

   As we view against the direction of propagation $\hat{z}$, we see that the field vector $\boldsymbol{E}(t)$ rotates *counterclockwise* with an angular frequency ω. The tip of this vector describes a circle. This is shown in Fig. 1.6(a). This left-circular polarization is also called *positive helicity*. Its eigenvector is

   $$\hat{e}_+ \equiv \frac{\hat{x} + \mathrm{i}\hat{y}}{\sqrt{2}}. \tag{1.72}$$

   b. **Right-circular polarization**. For $\varphi = -\pi/2$, also $\varepsilon = -\pi/4$, the wave is *right-circularly polarized* if it propagates in the positive z direction. We then have

   $$\boldsymbol{\mathcal{E}} = \mathcal{E}\frac{\hat{x} - \mathrm{i}\hat{y}}{\sqrt{2}} = \mathcal{E}\hat{e}_-, \tag{1.73}$$

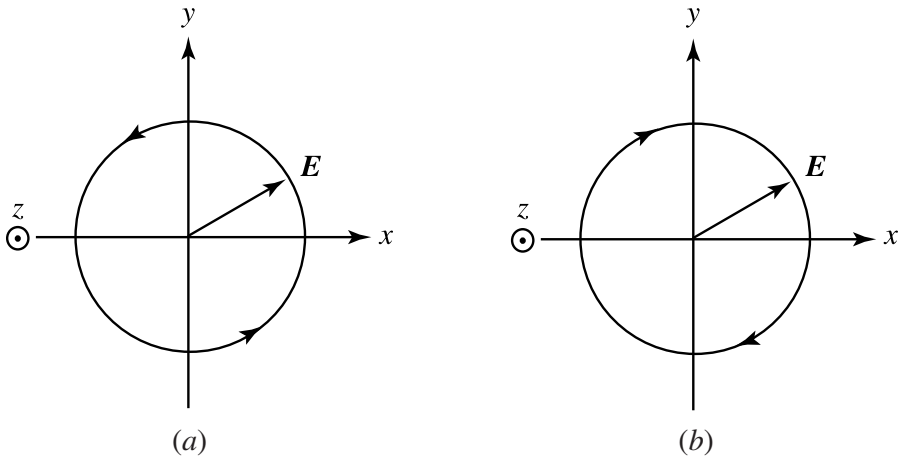


Figure 1.6 (*a*) Field of a left-circularly polarized wave. (*b*) Field of a right-circularly polarized wave.

and

$$\boldsymbol{E}(t) = \sqrt{2}\mathcal{E}(\hat{x}\cos\omega t - \hat{y}\sin\omega t). \tag{1.74}$$

The tip of this field vector rotates *clockwise* in a circle, as shown in Fig. 1.6(*b*). This right-circular polarization is also called *negative helicity*. Its eigenvector is

$$\hat{e}_- \equiv \frac{\hat{x} - \mathrm{i}\hat{y}}{\sqrt{2}}. \tag{1.75}$$

As can be seen, neither $\hat{e}_+$ nor $\hat{e}_-$ is a real vector. Note that the identification of $\hat{e}_+$, defined in (1.72), with left-circular polarization and that of $\hat{e}_-$, defined in (1.75), with right-circular polarization are based on the assumption that the wave propagates in the positive z direction. For a wave that propagates in the negative z direction, the handedness of these unit vectors changes: $\hat{e}_+$ becomes right-circular polarization, while $\hat{e}_-$ becomes left-circular polarization.

Linearly polarized light can be produced from unpolarized light using a *polarizer*. A polarizer can be of *transmission type*, which often utilizes the phenomenon of double refraction in an anisotropic crystal, discussed in Section 1.6, or of *reflection type*, which takes advantage of the polarization-sensitive reflectivity of a surface, discussed in Section 1.7. A very convenient transmission-type polarizer is the *Polaroid* film, which utilizes a material with *linear dichroism*, having low absorption for light linearly polarized in a particular direction and high absorption for light polarized orthogonally to this direction. The output is linearly polarized in the direction defined by the polarizer irrespective of the polarization state of the input optical wave. A polarizer can also be used to analyze the polarization of a particular optical wave. When so used, a polarizer is also called an *analyzer*.

1.5 Propagation in an isotropic medium

The propagation of an optical wave is governed by the wave equation. It depends on the optical property and physical structure of the medium. It also depends on the makeup of the optical wave, such as its frequency contents and its temporal characteristics. In this section, we consider the basic characteristics of the propagation of a monochromatic plane optical wave in an infinite homogeneous medium. For such a monochromatic wave, there is only one value of $\mathbf{k}$ and one value of ω. Its complex electric field is that given by (1.58), in which the field amplitude $\mathcal{E}$ is independent of $\mathbf{r}$ and t. Thus,

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \boldsymbol{\chi}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{r}, t) \tag{1.76}$$

and

$$\mathbf{D}(\mathbf{r}, t) = \boldsymbol{\epsilon}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{r}, t). \tag{1.77}$$

Also, in this section, we shall assume no spatial nonlocality in the media thus neglecting the $\mathbf{k}$ dependence of $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$. Then,

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \boldsymbol{\chi}(\omega) \cdot \mathbf{E}(\mathbf{r}, t) \tag{1.78}$$

and

$$\mathbf{D}(\mathbf{r}, t) = \boldsymbol{\epsilon}(\omega) \cdot \mathbf{E}(\mathbf{r}, t). \tag{1.79}$$

For a monochromatic wave of a frequency ω, the wave equation is simply

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E} + \mu_0 \boldsymbol{\epsilon}(\omega) \cdot \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0. \tag{1.80}$$

For an isotropic medium, $\boldsymbol{\epsilon}(\omega)$ is reduced to a scalar $\epsilon(\omega)$ and

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon(\omega)} \boldsymbol{\nabla} \cdot \mathbf{D} = 0. \tag{1.81}$$

Then, by using the vector identity $\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times = \boldsymbol{\nabla}\boldsymbol{\nabla} \cdot -\nabla^2$, the wave equation in (1.80) is reduced to the following simple form:

$$\nabla^2 \mathbf{E} - \mu_0 \epsilon(\omega) \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0. \tag{1.82}$$

For an anisotropic medium, (1.82) is generally not valid because (1.81) does not hold.

Note that with $\mathcal{E}$ in (1.58) being independent of $\mathbf{r}$ and t, we can make the following replacement for the operators when operating on $\mathbf{E}$ of the form in (1.58) or $\mathbf{H}$ of the same form:

$$\boldsymbol{\nabla} \longrightarrow \mathrm{i}\mathbf{k}, \qquad \frac{\partial}{\partial t} \longrightarrow -\mathrm{i}\omega. \tag{1.83}$$

Free space

In free space, $\mathbf{P} = 0$ and $\boldsymbol{\epsilon}$ is reduced to the scalar ϵ_0. Substitution of (1.58) in (1.82) then yields

$$k^2 = \omega^2 \mu_0 \epsilon_0. \tag{1.84}$$

The *propagation constant* in free space is

$$k = \frac{\omega}{c} = \frac{2\pi\nu}{c} = \frac{2\pi}{\lambda}, \tag{1.85}$$

where ν is the frequency of the optical wave and λ is its wavelength. Because k is proportional to $1/\lambda$, it is also called the *wavenumber*.

Using (1.83) and noting that $\mathbf{B} = \mu_0 \mathbf{H}$ and $\mathbf{D} = \epsilon_0 \mathbf{E}$, Maxwell's equations in (1.40)–(1.43) become

$$\mathbf{k} \times \mathbf{E} = \omega \mu_0 \mathbf{H}, \tag{1.86}$$

$$\mathbf{k} \times \mathbf{H} = -\omega \epsilon_0 \mathbf{E}, \tag{1.87}$$

$$\mathbf{k} \cdot \mathbf{E} = 0, \tag{1.88}$$

$$\mathbf{k} \cdot \mathbf{H} = 0. \tag{1.89}$$

From (1.86) and (1.87), we also have

$$\mathbf{E} \cdot \mathbf{H} = 0. \tag{1.90}$$

Therefore, the three vectors **E**, **H**, and **k** are orthogonal. These relationships also imply that

$$\mathbf{S} \parallel \mathbf{k}. \tag{1.91}$$

The relationships among the directions of these vectors are shown in Fig. 1.7.

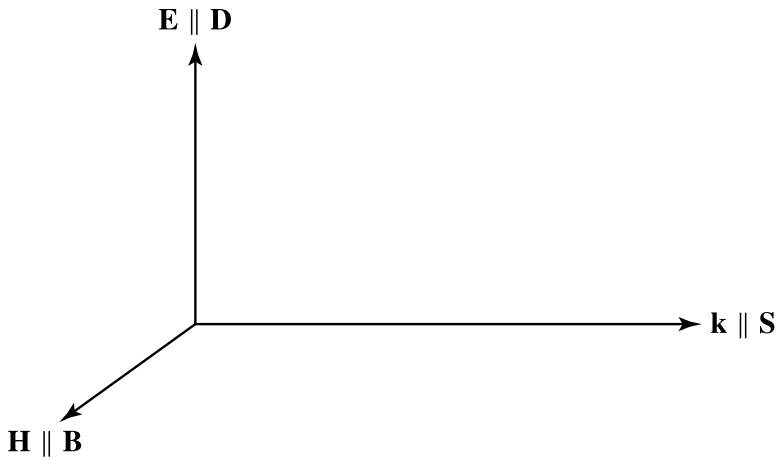


Figure 1.7 Relationships among the directions of **E**, **D**, **H**, **B**, **k**, and **S** in free space or in an isotropic medium.

Using (1.85), we can also write (1.86) and (1.87) in the following form:

$$\mathbf{H} = \frac{1}{Z_0}\hat{k} \times \mathbf{E}, \qquad \mathbf{E} = Z_0\mathbf{H} \times \hat{k}, \tag{1.92}$$

where $\hat{k} = \mathbf{k}/k$ is the unit vector in the $\mathbf{k}$ direction and

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 120\pi\ \Omega \approx 377\ \Omega \tag{1.93}$$

is the free-space *impedance*. The concept of this impedance is not that of the impedance of a resistor but is analogous to the concept of the impedance of a tranmission line.

Because $\mathbf{S} \parallel \mathbf{k}$, the light intensity in free space can be expressed as

$$I = \hat{k} \cdot \overline{\boldsymbol{S}} = 2\frac{|\mathbf{E}|^2}{Z_0} = 2Z_0|\mathbf{H}|^2. \tag{1.94}$$

Lossless medium

In this case, $\boldsymbol{\epsilon}(\omega)$ is reduced to a positive real scalar $\epsilon(\omega)$, which is different from ϵ_0. All of the results obtained for free space remain valid, except that ϵ_0 is replaced by $\epsilon(\omega)$. This change of the electric permittivity from a vacuum to a material is measured by the *relative electric permittivity*, ϵ/ϵ_0, which is a dimensionless quantity also known as the *dielectric constant* of the material. Therefore, the propagation constant in the medium is

$$k = \omega\sqrt{\mu_0\epsilon} = \frac{n\omega}{c} = \frac{2\pi n\nu}{c} = \frac{2\pi n}{\lambda}, \tag{1.95}$$

where

$$n = \sqrt{\frac{\epsilon}{\epsilon_0}} = (\text{dielectric constant})^{1/2} \tag{1.96}$$

is the *index of refraction*, or *refractive index*, of the medium.

In a medium that has an index of refraction n, the optical frequency is still ν, but the optical wavelength is λ/n, and the speed of light is $v = c/n$. Because $n(\omega)$ in a medium is generally frequency dependent, the speed of light in a medium is also frequency dependent. This results in various dispersive phenomena such as the separation of different colors by a prism and the broadening or shortening of an optical pulse traveling through a medium. We also find that

$$Z = \frac{Z_0}{n} \tag{1.97}$$

in a medium. The light intensity is then

$$I = 2\frac{|\mathbf{E}|^2}{Z} = 2Z|\mathbf{H}|^2 = \frac{2k}{\omega\mu_0}|\mathbf{E}|^2 = \frac{2k}{\omega\epsilon}|\mathbf{H}|^2. \tag{1.98}$$

Medium with a loss or gain

As discussed in the preceding section, χ and ϵ become complex when a medium has an optical loss or gain. Therefore,

$$k^2 = \omega^2 \mu_0 \epsilon = \omega^2 \mu_0 (\epsilon' + \mathrm{i}\epsilon''), \tag{1.99}$$

and the propagation constant k becomes complex:

$$k = k' + \mathrm{i}k'' = \beta + \mathrm{i}\frac{\alpha}{2}. \tag{1.100}$$

The index of refraction also becomes complex:

$$n = \sqrt{\frac{\epsilon' + \mathrm{i}\epsilon''}{\epsilon_0}} = n' + \mathrm{i}n''. \tag{1.101}$$

The relation between k and n in (1.95) is still valid. Meanwhile, the impedance Z of the medium also becomes complex. Therefore, $\mathbf{E}$ and $\mathbf{H}$ are no longer in phase, as can be seen from (1.92) by replacing Z_0 with a complex Z, and I is not simply given by (1.98) but is given by the real part of it.

It can be shown that if we choose β to be positive, the sign of α is the same as that of ϵ''. In this case, k' and n' are also positive and k'' and n'' also have the same sign as ϵ''. If we consider as an example an optical wave propagating in the z direction, then $\hat{k} = \hat{z}$ and, from (1.58) and (1.100), the complex electric field is

$$\mathbf{E}(\mathbf{r}, t) = \boldsymbol{\mathcal{E}} \mathrm{e}^{-\alpha z/2} \exp(\mathrm{i}\beta z - \mathrm{i}\omega t). \tag{1.102}$$

It can be seen that the wave has a phase that varies sinusoidally with a period of $1/\beta$ along z. In addition, its amplitude is not constant but varies exponentially with z. Thus, light intensity is also a function of z:

$$I \propto \mathrm{e}^{-\alpha z}. \tag{1.103}$$

Clearly, β is the *wavenumber* in this case, and the sign of α determines the attenuation or amplification of the optical wave:

1. If $\chi'' > 0$, then $\epsilon'' > 0$ and $\alpha > 0$. As the optical wave propagates, its field amplitude and intensity decay exponentially along the direction of propagation. Therefore, α is called the *absorption coefficient* or *attenuation coefficient*.
2. If $\chi'' < 0$, then $\epsilon'' < 0$ and $\alpha < 0$. The field amplitude and intensity of the optical wave grow exponentially. Then, we define $g = -\alpha$ as the *gain coefficient* or *amplification coefficient*.

The unit of both α and g is per meter, often also quoted per centimeter.

EXAMPLE 1.1 The complex susceptibility of GaAs at an optical wavelength of $\lambda = 850\,\text{nm}$ is $\chi = 12.17 + \text{i}0.49$. Therefore, at this wavelength, GaAs has a complex refractive index of

$$n = (\epsilon/\epsilon_0)^{1/2} = (1+\chi)^{1/2} = (13.17 + \text{i}0.49)^{1/2} = 3.63 + \text{i}0.0676$$

and an absorption coefficient of

$$\alpha = 2k'' = \frac{4\pi n''}{\lambda} = \frac{4\pi \times 0.0676}{850 \times 10^{-9}}\,\text{m}^{-1} = 10^6\,\text{m}^{-1}.$$

An optical beam at 850 nm wavelength can travel in GaAs only for a distance of $l = -\ln(1-0.99)/\alpha = 4.6\,\mu\text{m}$ before losing 99% of its energy to absorption, which is obtained by solving $1 - \text{e}^{-\alpha l} = 0.99$ with $\alpha = 10^6\,\text{m}^{-1}$.

1.6 Propagation in an anisotropic medium

In an anisotropic medium, the tensors $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ do not reduce to scalars. Therefore, $\mathbf{P} \nparallel \mathbf{E}$ and $\mathbf{D} \nparallel \mathbf{E}$. As a result, (1.81) is not true any more, and, in general,

$$\boldsymbol{\nabla} \cdot \mathbf{E} \neq 0. \tag{1.104}$$

Consequently, (1.82) cannot be used for propagation of a monochromatic wave in an anisotropic medium. Instead, (1.80) has to be used.

Anisotropic $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$

In a linear anisotropic medium, both $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ are second-rank tensors. They can be expressed in the following matrix forms:

$$\boldsymbol{\chi} = \begin{bmatrix} \chi_{11} & \chi_{12} & \chi_{13} \\ \chi_{21} & \chi_{22} & \chi_{23} \\ \chi_{31} & \chi_{32} & \chi_{33} \end{bmatrix} \tag{1.105}$$

and

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix}. \tag{1.106}$$

The relationships $\mathbf{P} = \epsilon_0 \boldsymbol{\chi} \cdot \mathbf{E}$ and $\mathbf{D} = \boldsymbol{\epsilon} \cdot \mathbf{E}$ are carried out as products between a tensor and a column vector. For example,

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}. \tag{1.107}$$

In general, the matrix in (1.106) representing the tensor $\boldsymbol{\epsilon}$ is not diagonal. It can be diagonalized by a proper choice of the coordinate system, yielding

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}, \tag{1.108}$$

where ϵ_i, for $i = 1, 2, 3$, are the *eigenvalues* of $\boldsymbol{\epsilon}$ with their corresponding *eigenvectors*, $\hat{u}_i$, being the axes of the coordinate system chosen to diagonalize $\boldsymbol{\epsilon}$. The characteristics of ϵ_i and $\hat{u}_i$ depend on the symmetry properties of $\boldsymbol{\epsilon}$. The two matrices representing $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ have the same symmetry properties because $\boldsymbol{\epsilon} = \epsilon_0(1 + \boldsymbol{\chi})$, where 1 has the form of a 3×3 identity matrix in its addition to the tensor $\boldsymbol{\chi}$. Therefore, $\boldsymbol{\chi}$ and $\boldsymbol{\epsilon}$ are diagonalized by the same set of eigenvectors that represent the axes of the chosen coordinate system.

The symmetry properties of $\boldsymbol{\epsilon}$, as well as those of $\boldsymbol{\chi}$, are determined by the properties of the medium.

1. **Reciprocal media.** Nonmagnetic materials in the absence of an external magnetic field are *reciprocal media*. In a reciprocal medium, the *Lorentz reciprocity theorem* of electromagnetics holds; consequently, the source and the detector of an optical signal can be interchanged. If such a material is not *optically active*, its optical properties are described by a symmetric $\boldsymbol{\epsilon}$ tensor: $\epsilon_{ij} = \epsilon_{ji}$. *For a symmetric tensor, the eigenvectors $\hat{u}_i$ are always real vectors.* They can be chosen to be $\hat{x}$, $\hat{y}$, and $\hat{z}$ of a rectangular coordinate system in real space. This is true even when $\boldsymbol{\epsilon}$ is complex. (a) If a nonmagnetic medium does not have an optical loss or gain, its $\boldsymbol{\epsilon}$ tensor is Hermitian. A symmetric Hermitian tensor is real and symmetric: $\epsilon_{ij}^* = \epsilon_{ij} = \epsilon_{ji} = \epsilon_{ji}^*$. In this case, the eigenvalues ϵ_i have real values. (b) If a nonmagnetic medium has an optical loss or gain, its $\boldsymbol{\epsilon}$ tensor is not Hermitian but is complex and symmetric: $\epsilon_{ij} = \epsilon_{ji}$ but $\epsilon_{ij} \neq \epsilon_{ji}^*$. Then, the eigenvalues ϵ_i are complex. (c) If a nonmagnetic medium is optically active, it is still reciprocal although its $\boldsymbol{\epsilon}$ tensor is not symmetric. In this case, the eigenvectors are complex but the eigenvalues can be real if the medium is lossless.
2. **Nonreciprocal media.** Magnetic materials, and nonmagnetic materials subject to an external magnetic field, are *nonreciprocal media*. In such a medium, no symmetry exists when the source and the detector of an optical signal are interchanged. The $\boldsymbol{\epsilon}$ tensor describing the optical properties of such a material is not symmetric: $\epsilon_{ij} \neq \epsilon_{ji}$. *The eigenvectors $\hat{u}_i$ are complex vectors.* Therefore, they are not ordinary coordinate axes in real space, as seen later in the discussion on magneto-optic devices. (a) For a lossless magnetic medium, $\boldsymbol{\epsilon}$ is Hermitian: $\epsilon_{ij} = \epsilon_{ji}^*$. In this case, the eigenvalues ϵ_i are real even though the eigenvectors are complex. (b) For a magnetic medium

that has an optical loss or gain, $\boldsymbol{\epsilon}$ is neither symmetric nor Hermitian. Both the eigenvectors and the eigenvalues are complex.

Most materials used for photonic devices are nonmagnetic dielectric materials that are not optically active. The properties of magnetic materials are of interest to us only in consideration of magneto-optic devices, discussed in Chapter 7. Similarities and differences between magnetic and optically active materials are also briefly mentioned in Section 7.2. The discussion in the rest of this section is specific to nonmagnetic dielectric materials that are not optically active.

According to the above, in a dielectric material the axes of the coordinate system in which $\boldsymbol{\epsilon}$ is diagonal are real in space and can be labeled $\hat{x}$, $\hat{y}$, and $\hat{z}$. Noncrystalline materials are generally isotropic, for which the choice of the orthogonal coordinate axes $\hat{x}$, $\hat{y}$, and $\hat{z}$ is arbitrary. In contrast, many crystalline materials that are useful for photonic device applications are anisotropic. For any given anisotropic crystal, there is a unique set of coordinate axes for $\boldsymbol{\epsilon}$ to be diagonal. These unique $\hat{x}$, $\hat{y}$, and $\hat{z}$ coordinate axes are called the *principal dielectric axes*, or simply the *principal axes*, of the crystal. In the coordinate system defined by these principal axes, $\boldsymbol{\epsilon}$ is diagonalized with eigenvalues ϵ_x, ϵ_y, and ϵ_z. The components of $\mathbf{D}$ and $\mathbf{E}$ along these axes have the following simple relations:

$$D_x = \epsilon_x E_x, \qquad D_y = \epsilon_y E_y, \qquad D_z = \epsilon_z E_z. \tag{1.109}$$

The values ϵ_x/ϵ_0, ϵ_y/ϵ_0, and ϵ_z/ϵ_0 are the eigenvalues of the *dielectric constant tensor*, $\boldsymbol{\epsilon}/\epsilon_0$, and are called the *principal dielectric constants*. They define three *principal indices of refraction*:

$$n_x = \sqrt{\frac{\epsilon_x}{\epsilon_0}}, \qquad n_y = \sqrt{\frac{\epsilon_y}{\epsilon_0}}, \qquad n_z = \sqrt{\frac{\epsilon_z}{\epsilon_0}}. \tag{1.110}$$

Note that when $\boldsymbol{\epsilon}$ is diagonalized, $\boldsymbol{\chi}$ is also diagonalized along the same principal axes with corresponding *principal dielectric susceptibilities*, χ_x, χ_y, and χ_z. The principal dielectric susceptibilities of any lossless dielectric material always have positive values; therefore, the principal dielectric constants of such a material are always larger than unity.

Because $\mathbf{D} \perp \mathbf{k}$ due to the fact that $\boldsymbol{\nabla} \cdot \mathbf{D} = 0$, there is no $\mathbf{D}$ component along the direction of wave propagation. In general, $\mathbf{D}$ can be decomposed into two mutually orthogonal components, each of which is also orthogonal to $\mathbf{k}$. In an anisotropic crystal, these two components generally have different indices of refraction, and thus different propagation constants. This phenomenon is called *birefringence*. Such a crystal is a *birefringent crystal*.

EXAMPLE 1.2 At an optical wavelength of 1 μm, the permittivity tensor of the KDP crystal represented in a rectangular coordinate system defined by $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$ is found to be

$$\boldsymbol{\epsilon} = \epsilon_0 \begin{bmatrix} 2.28 & 0 & 0 \\ 0 & 2.25 & -0.051\,96 \\ 0 & -0.051\,96 & 2.19 \end{bmatrix}.$$

Find the principal axes and the corresponding principal indices for this crystal.

Solution Note that $\boldsymbol{\epsilon}$ is represented by a symmetric matrix because KDP is a nonmagnetic dielectric crystal. Diagonalization of this matrix yields the following eigenvalues and corresponding eigenvectors:

$$\epsilon_x = 2.28\epsilon_0,\ \hat{x} = \hat{x}_1,$$
$$\epsilon_y = 2.28\epsilon_0,\ \hat{y} = 0.866\hat{x}_2 - 0.500\hat{x}_3,$$
$$\epsilon_z = 2.16\epsilon_0,\ \hat{z} = 0.500\hat{x}_2 + 0.866\hat{x}_3.$$

Therefore, the principal axes of the crystal are $\hat{x}$, $\hat{y}$, and $\hat{z}$, given above, and the principal indices of refraction are $n_x = \sqrt{2.28} = 1.51$, $n_y = \sqrt{2.28} = 1.51$, and $n_z = \sqrt{2.16} = 1.47$.

Index ellipsoid

The inverse of the dielectric constant tensor mentioned above is the *relative impermeability tensor*:

$$\boldsymbol{\eta} = [\eta_{ij}] = \left(\frac{\boldsymbol{\epsilon}}{\epsilon_0}\right)^{-1}, \tag{1.111}$$

where i and j are spatial coordinate indices. In a general rectangular coordinate system (x_1, x_2, x_3), the ellipsoid defined by

$$\sum_{i,j} x_i \eta_{ij} x_j = 1 \tag{1.112}$$

is called the *index ellipsoid* or the *optical indicatrix*. In a nonmagnetic dielectric medium, $\boldsymbol{\eta}$ is a symmetric tensor, i.e., $\eta_{ij} = \eta_{ji}$, because $\boldsymbol{\epsilon}$ is symmetric. Therefore, (1.112) can be written as

$$\eta_{11}x_1^2 + \eta_{22}x_2^2 + \eta_{33}x_3^2 + 2\eta_{23}x_2x_3 + 2\eta_{31}x_3x_1 + 2\eta_{12}x_1x_2 = 1. \tag{1.113}$$

This equation is usually written as

$$\eta_1 x_1^2 + \eta_2 x_2^2 + \eta_3 x_3^2 + 2\eta_4 x_2 x_3 + 2\eta_5 x_3 x_1 + 2\eta_6 x_1 x_2 = 1 \tag{1.114}$$

using the following *index contraction* rule to reduce the double index ij of η_{ij} to the

single index α of η_α:

$$\begin{array}{lcccccc} ij: & 11 & 22 & 33 & 23, 32 & 31, 13 & 12, 21 \\ \text{or } ij: & xx & yy & zz & yz, zy & zx, xz & xy, yx \\ \alpha: & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \tag{1.115}$$

The index ellipsoid equation is invariant with respect to coordinate rotation. When a coordinate system with axes aligned with the principal dielectric axes of the crystal is chosen, ϵ is diagonalized. Thus the tensor $\boldsymbol{\eta}$ is also diagonalized with the following eigenvalues:

$$\eta_x = \frac{\epsilon_0}{\epsilon_x} = \frac{1}{n_x^2}, \qquad \eta_y = \frac{\epsilon_0}{\epsilon_y} = \frac{1}{n_y^2}, \qquad \eta_z = \frac{\epsilon_0}{\epsilon_z} = \frac{1}{n_z^2}. \tag{1.116}$$

In this coordinate system, the index ellipsoid takes the following simple form:

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1. \tag{1.117}$$

Comparing (1.117) with (1.114), we find that the terms containing cross products of different coordinates are eliminated when the coordinate system of the principal dielectric axes is used. The principal axes of the index ellipsoid now coincide with the principal dielectric axes of the crystal, and the principal indices of refraction of the crystal are given by the semiaxes of the index ellipsoid. This is illustrated in Fig. 1.8. Therefore, a coordinate transformation by rotation to eliminate cross-product terms in the index ellipsoid equation is equivalent to diagonalization of the ϵ tensor. The

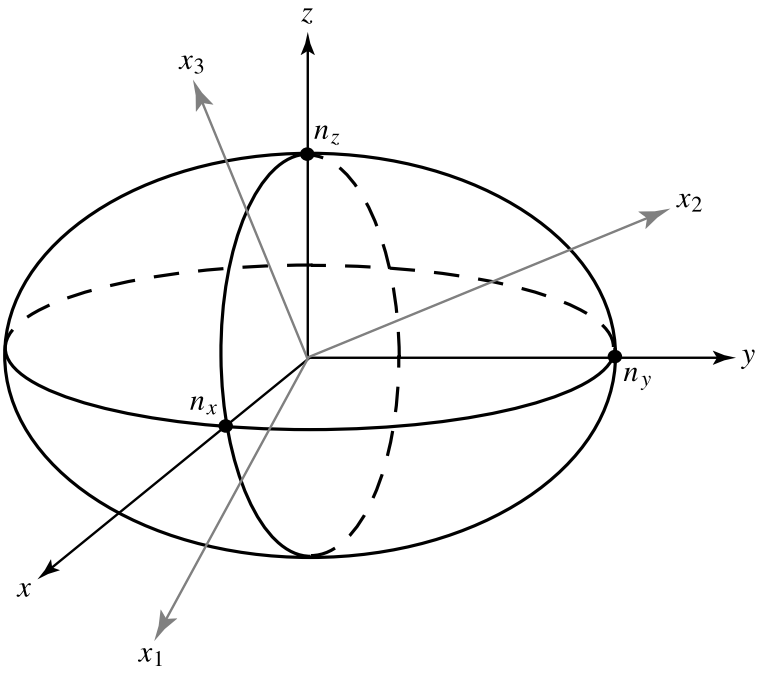


Figure 1.8 Index ellipsoid and its relationship with the coordinate system. Here (x, y, z) is the coordinate system aligned with the principal axes of the crystal, while (x_1, x_2, x_3) is an arbitrary coordinate system.

principal dielectric axes and their corresponding principal indices of refraction can be found through either approach. Between the two approaches, however, diagonalization of the ϵ tensor is better because it is more systematic and is easier to carry out.

EXAMPLE 1.3 Find the principal axes and their corresponding principal indices for the KDP crystal given in Example 1.2 by using the index ellipsoid instead of diagonalizing the ϵ tensor as done in Example 1.2. Compare the two approaches.

Solution The relative impermeability tensor in the (x_1, x_2, x_3) coordinate system can be found by inverting the ϵ tensor:

$$\eta = \left(\frac{\epsilon}{\epsilon_0}\right)^{-1} = \begin{bmatrix} 2.28 & 0 & 0 \\ 0 & 2.25 & -0.051\,96 \\ 0 & -0.051\,96 & 2.19 \end{bmatrix}^{-1} \approx \begin{bmatrix} \frac{1}{2.28} & 0 & 0 \\ 0 & \frac{1}{2.25} & 0.010\,55 \\ 0 & 0.010\,55 & \frac{1}{2.19} \end{bmatrix}.$$

In the (x_1, x_2, x_3) coordinate system, the index ellipsoid is thus described by the following equation:

$$\frac{x_1^2}{2.28} + \frac{x_2^2}{2.25} + \frac{x_3^2}{2.19} + 0.0211 x_2 x_3 = 1.$$

To find the principal axes and their principal indices of refraction, the cross-product term has to be eliminated by rotating the coordinates. From Example 1.2, we know that this can be done by taking

$$x_1 = x, \quad x_2 = 0.866y + 0.500z, \quad x_3 = -0.500y + 0.866z.$$

Substitution of these relations into the above index ellipsoid equation transforms it into the following equation for the index ellipsoid in the (x, y, z) coordinate system:

$$\frac{x^2}{2.28} + \frac{y^2}{2.28} + \frac{z^2}{2.16} = 1.$$

Thus the principal indices are $n_x = \sqrt{2.28} = 1.51$, $n_y = \sqrt{2.28} = 1.51$, and $n_z = \sqrt{2.16} = 1.47$.

Comparing the two approaches illustrated in this example and in Example 1.2, it is clear that they are equivalent to one another. It is also clear that the method of diagonalizing ϵ described in Example 1.2 is more systematic and straightforward than that of eliminating the cross-product terms in the equation for the index ellipsoid, particularly when there is more than one cross-product term.

Propagation along a principal axis

We first consider the simple case when an optical wave propagates along one of the principal axes, say $\hat{z}$. Then the field can be decomposed into two *normal modes*, each of which is polarized along one of the other two principal axes, $\hat{x}$ or $\hat{y}$. We see from (1.109) and (1.110) that each field component along a principal axis has a characteristic index of refraction n_i, meaning that it has a characteristic propagation constant of $k^i = n_i\omega/c$, which is determined by the polarization of the field but not by the direction of wave propagation. For a wave propagating along $\hat{z}$, the electric field can be expressed as

$$\begin{aligned}\mathbf{E} &= \hat{x}\mathcal{E}_x e^{ik^x z - i\omega t} + \hat{y}\mathcal{E}_y e^{ik^y z - i\omega t} \\ &= \left[\hat{x}\mathcal{E}_x + \hat{y}\mathcal{E}_y e^{i(k^y - k^x)z}\right] e^{ik^x z - i\omega t}.\end{aligned} \tag{1.118}$$

Because the wave propagates in the z direction, the wavevectors are $\mathbf{k}^x = k^x\hat{z}$ for the x-polarized field and $\mathbf{k}^y = k^y\hat{z}$ for the y-polarized field. Note that $k^x = n_x\omega/c$ and $k^y = n_y\omega/c$ are the propagation constants of the x- and y-polarized fields, respectively, not to be confused with the x and y components of a wavevector $\mathbf{k}$, which are normally expressed as k_x and k_y. The field expressed in (1.118) has the following propagation characteristics.

1. If it is originally linearly polarized along one of the principal axes, it remains linearly polarized in the same direction.
2. If it is originally linearly polarized at an angle $\theta = \tan^{-1}(\mathcal{E}_y/\mathcal{E}_x)$ with respect to the x axis, its polarization state varies periodically along z with a period of $2\pi/|k^y - k^x|$. In general, its polarization follows a sequence of variations from linear to elliptical to linear in the first half-period and then reverses the sequence back to linear in the second half-period. At the half-period position, it is linearly polarized at an angle θ on the other side of the x axis. Thus the polarization is rotated by 2θ from the original direction. This is shown in Fig. 1.9(a). In the special case when $\theta = 45^\circ$, the wave is circularly polarized at the quarter-period point and is linearly polarized with its polarization rotated by 90° from the original direction at the half-period point. This is shown in Fig. 1.9(b).

These characteristics have very useful applications. A plate of an anisotropic material that has a quarter-period thickness of

$$l_{\lambda/4} = \frac{1}{4}\cdot\frac{2\pi}{|k^y - k^x|} = \frac{\lambda}{4|n_y - n_x|} \tag{1.119}$$

is called a *quarter-wave plate*. It can be used to convert a linearly polarized wave to circular or elliptic polarization, and vice versa. A plate of thickness $3l_{\lambda/4}$ or $5l_{\lambda/4}$ or any odd integral multiple of $l_{\lambda/4}$ also has the same function. In contrast, a plate of a

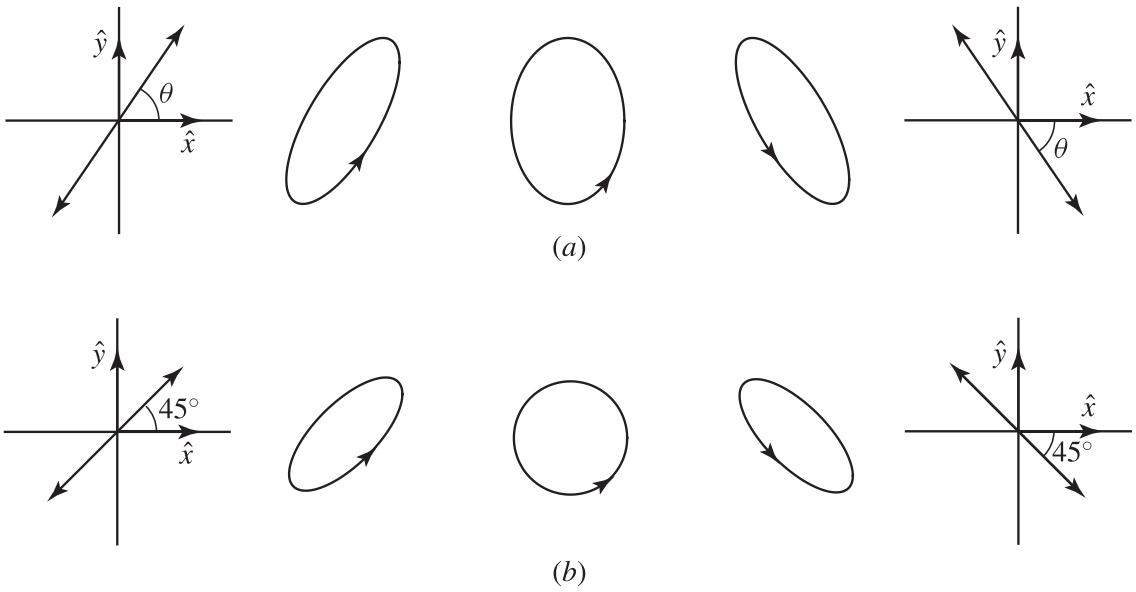


Figure 1.9 Evolution of the polarization state of an optical wave propagating along the principal axis $\hat{z}$ of an anisotropic crystal that has $n_x \neq n_y$. Only the evolution over one half-period is shown here. (a) The optical wave is initially linearly polarized at an arbitrary angle θ with respect to the principal axis $\hat{x}$. (b) The optical wave is initially polarized at 45° with respect to $\hat{x}$.

half-period thickness of

$$l_{\lambda/2} = \frac{\lambda}{2|n_y - n_x|} \tag{1.120}$$

is called a *half-wave plate*. It can be used to rotate the polarization direction of a linearly polarized wave by any angular amount by properly choosing the angle θ between the incident polarization with respect to the principal axis $\hat{x}$, or $\hat{y}$, of the crystal. A plate of a thickness that is any odd integral multiple of $l_{\lambda/2}$ has the same function. Note that though the output from a quarter-wave or half-wave plate can be linearly polarized, the wave plates are not polarizers. They are based on different principles and have completely different functions.

For the quarter-wave and half-wave plates discussed here, $n_x \neq n_y$. Between the two crystal axes $\hat{x}$ and $\hat{y}$, the one with the smaller index is called the *fast axis* while the other, with the larger index, is the *slow axis*.

EXAMPLE 1.4 KDP can be used to make quarter-wave and half-wave plates. Find the thicknesses of the quarter-wave and half-wave plates made of KDP for 1 μm wavelength.

Solution From Example 1.3, we know that $n_x = n_y = 1.51$ and $n_z = 1.47$ for KDP at 1 μm wavelength. Because $n_x = n_y$, we cannot use n_x and n_y to make a wave plate that allows the beam to propagate in the z direction. Instead, the beam can propagate in any direction on the xy plane so that the difference between n_z and $n_x = n_y$ can

be used for the function of a wave plate. Assuming that the wave propagates in the x direction, then the thickness of a quarter-wave plate for $\lambda = 1$ μm is

$$l_{\lambda/4} = \frac{\lambda}{4|n_y - n_z|} = \frac{1\ \mu\text{m}}{4 \times |1.51 - 1.47|} = 6.25\ \mu\text{m}.$$

A quarter-wave plate at 1 μm wavelength can have a thickness of any odd integral multiple, such as 18.75 μm, 31.25 μm, . . . , of 6.25 μm. A half-wave plate for the 1 μm wavelength has a thickness of

$$l_{\lambda/2} = \frac{\lambda}{2|n_y - n_z|} = \frac{1\ \mu\text{m}}{2 \times |1.51 - 1.47|} = 12.5\ \mu\text{m}.$$

A plate of a thickness that is an odd multiple, such as 37.5 μm, 62.5 μm, . . . , of 12.5 μm also functions as a half-wave plate at 1 μm wavelength. For these wave plates, $\hat{z}$ is the fast axis and $\hat{y}$ is the slow axis because $n_z < n_y$.

Optical axes

The state of polarization of an optical wave generally varies along its path of propagation through an anisotropic crystal unless it is linearly polarized in the direction of a principal axis. However, in an anisotropic crystal with $n_x = n_y \neq n_z$, a wave propagating in the z direction does not see the anisotropy of the crystal because in this situation the x and y components of the field have the same propagation constant. This wave will maintain its original polarization as it propagates through the crystal. Evidently, this is true only for propagation along the z axis in such a crystal. Such a unique axis in a crystal along which an optical wave can propagate with an index of refraction that is independent of its polarization direction is called the *optical axis* of the crystal.

For an anisotropic crystal that has only one distinctive principal index among its three principal indices, there is only one optical axis, which coincides with the axis of the distinctive principal index of refraction. Such a crystal is called a *uniaxial crystal*. It is customary to assign $\hat{z}$ to this unique principal axis. The identical principal indices of refraction are called the *ordinary index*, n_o, and the distinctive index of refraction is called the *extraordinary index*, n_e. Thus, $n_x = n_y = n_\text{o}$ and $n_z = n_\text{e}$. The crystal is called *positive uniaxial* if $n_\text{e} > n_\text{o}$ and is *negative uniaxial* if $n_\text{e} < n_\text{o}$.

For a crystal that has three distinct principal indices of refraction, there are two optical axes, neither of which coincides with any one of the principal axes. Such a crystal is called a *biaxial crystal* because of the existence of two optical axes.

Ordinary and extraordinary waves

When an optical wave propagates in a direction other than that along an optical axis, the index of refraction depends on the direction of its polarization. In this situation, there exist two normal modes of linearly polarized waves, each of which sees a unique

index of refraction. One of them is the polarization perpendicular to the optical axis. This normal mode is called the *ordinary wave*. We use $\hat{e}_o$ to indicate its direction of polarization. The other normal mode is clearly one that is perpendicular to $\hat{e}_o$ because the two normal-mode polarizations are orthogonal to each other. This normal mode is called the *extraordinary wave*, and its direction of polarization is indicated by $\hat{e}_e$. *Note that these are the directions of* **D** *rather than those of* **E**. For the ordinary wave, $\hat{e}_o \parallel \mathbf{D}_o \parallel \mathbf{E}_o$. For the extraordinary wave, $\hat{e}_e \parallel \mathbf{D}_e \not\parallel \mathbf{E}_e$ except when $\hat{e}_e$ is parallel to a principal axis. Both $\hat{e}_o$ and $\hat{e}_e$, being the unit vectors of $\mathbf{D}_o$ and $\mathbf{D}_e$, are perpendicular to the direction of wave propagation, $\hat{k}$. From this understanding, both $\hat{e}_o$ and $\hat{e}_e$ can be found if both $\hat{k}$ and the optical axis are known. For a uniaxial crystal with optical axis $\hat{z}$, this means that

$$\hat{e}_o = \frac{1}{\sin\theta}\hat{k} \times \hat{z}, \qquad \hat{e}_e = \hat{e}_o \times \hat{k} \tag{1.121}$$

if the vector $\hat{k}$ is in a direction that is at an angle θ with respect to $\hat{z}$ and an angle ϕ with respect to the axis $\hat{x}$. Therefore, we have (see Problem 1.6.12)

$$\hat{k} = \hat{x}\sin\theta\cos\phi + \hat{y}\sin\theta\sin\phi + \hat{z}\cos\theta, \tag{1.122}$$

$$\hat{e}_o = \hat{x}\sin\phi - \hat{y}\cos\phi, \tag{1.123}$$

$$\hat{e}_e = -\hat{x}\cos\theta\cos\phi - \hat{y}\cos\theta\sin\phi + \hat{z}\sin\theta. \tag{1.124}$$

The relationships among these vectors are illustrated in Fig. 1.10.

The indices of refraction associated with the ordinary and extraordinary waves can be found by using the index ellipsoid given in (1.117), as is shown in Fig. 1.11. The

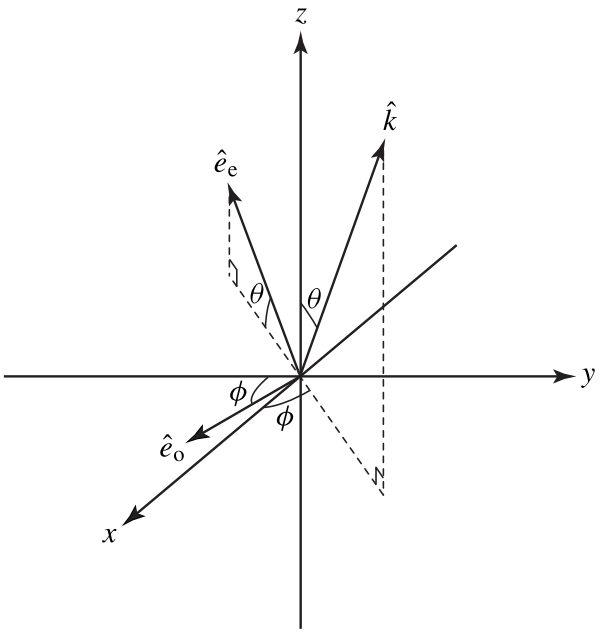


Figure 1.10 Relationships among the direction of wave propagation and the polarization directions of the ordinary and extraordinary waves.

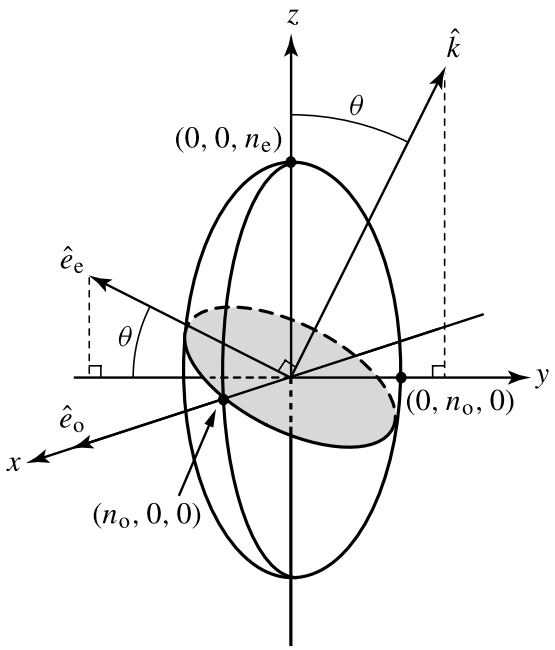


Figure 1.11 Determination of the indices of refraction for the ordinary and extraordinary waves in a uniaxial crystal using index ellipsoid.

intersection of the index ellipsoid and the plane normal to $\hat{k}$ at the origin of the ellipsoid defines an index ellipse. The principal axes of this index ellipse are in the directions of $\hat{e}_o$ and $\hat{e}_e$, and their half-lengths are the corresponding indices of refraction. For a uniaxial crystal, the index of refraction for the ordinary wave is simply n_o. The index of refraction for the extraordinary wave depends on the angle θ and is given by (see Problem 1.6.12)

$$\frac{1}{n_e^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2}, \tag{1.125}$$

which can be seen from Fig. 1.11. Because **D** is orthogonal to **k** and can be decomposed into $\mathbf{D}_o$ and $\mathbf{D}_e$ components, we have

$$\mathbf{D} = \hat{e}_o \mathcal{D}_o e^{ik_o \hat{k}\cdot\mathbf{r} - i\omega t} + \hat{e}_e \mathcal{D}_e e^{ik_e \hat{k}\cdot\mathbf{r} - i\omega t}, \tag{1.126}$$

where $k_o = n_o\omega/c$ and $k_e = n_e(\theta)\omega/c$. In general, **E** cannot be written in the form of (1.126) because its longitudinal component along the wave propagation direction **k** does not vanish except when $\theta = 0^\circ$ or 90°. We see that $n_e(0^\circ) = n_o$ and $n_e(90^\circ) = n_e$. The special case when the wave propagates along one of the principal axes discussed earlier belongs to one of these situations.

The normal-mode polarizations for an optical wave propagating in a biaxial crystal can be found following a similar, albeit more complicated, procedure.

EXAMPLE 1.5 From the preceding three examples, we find that KDP is a uniaxial crystal with $\hat{z}$ being its optical axis because $n_x = n_y \neq n_z$. At 1 μm wavelength, we have

$n_o = 1.51$ and $n_e = 1.47$. KDP is negative uniaxial because $n_o > n_e$. For an optical wave propagating in KDP along a direction $\hat{k}$ that makes an angle θ with respect to the optical axis $\hat{z}$, the refractive index for the extraordinary wave is a function of θ. For $\theta = 0^\circ, n_e(0^\circ) = n_o = 1.51$. For $\theta = 90^\circ, n_e(90^\circ) = n_e = 1.47$. For $0^\circ < \theta < 90^\circ$, $1.47 < n_e(\theta) < 1.51$. For example,

$$n_e(30^\circ) = \left(\frac{\cos^2 30^\circ}{n_o^2} + \frac{\sin^2 30^\circ}{n_e^2}\right)^{-1/2} = 1.50,$$

$$n_e(60^\circ) = \left(\frac{\cos^2 60^\circ}{n_o^2} + \frac{\sin^2 60^\circ}{n_e^2}\right)^{-1/2} = 1.48.$$

Spatial beam walk-off

Each of the normal modes has a well-defined propagation constant. Therefore, the fields of monochromatic ordinary and extraordinary waves in an anisotropic medium can be separately written in the form of (1.47), with $\mathbf{k} = \mathbf{k}_o$ for the ordinary way and $\mathbf{k} = \mathbf{k}_e$ for the extraordinary way. By using (1.83), Maxwell's equations for a normal mode, either ordinary or extraordinary, reduce to the following:

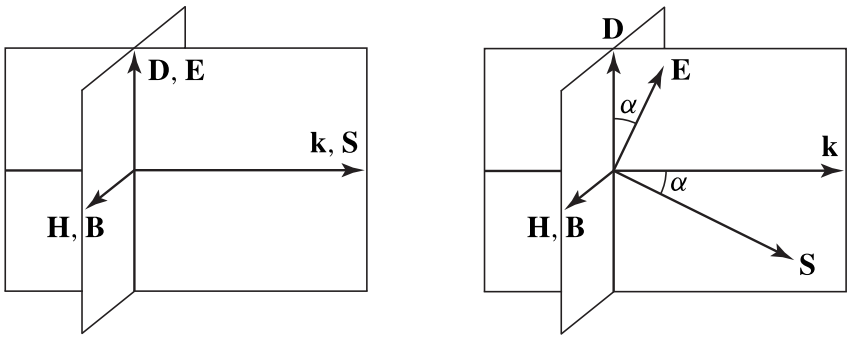
$$\mathbf{k} \times \mathbf{E} = \omega\mu_0\mathbf{H}, \tag{1.127}$$

$$\mathbf{k} \times \mathbf{H} = -\omega\mathbf{D}, \tag{1.128}$$

$$\mathbf{k} \cdot \mathbf{D} = 0, \tag{1.129}$$

$$\mathbf{k} \cdot \mathbf{H} = 0. \tag{1.130}$$

Note that because $n_o \neq n_e$, these relations apply to the ordinary and the extraordinary normal mode *separately* with different values for $\mathbf{k}$ but not to a wave mixing the two modes. At optical frequencies, $\mathbf{B} = \mu_0\mathbf{H}$ is also true in an anisotropic medium. Therefore, (1.127) and (1.130) have the same forms as (1.86) and (1.89), respectively. Because (1.88) for a wave in an isotropic medium is now replaced by (1.129), we have $\mathbf{D} \perp \mathbf{k}$ for both ordinary and extraordinary waves. For an ordinary wave, $\mathbf{E}_o \perp \mathbf{k}_o$ because $\mathbf{D}_o \parallel \mathbf{E}_o$. Therefore, the relationships shown in Fig. 1.12(*a*) among the field vectors for an ordinary wave in an anisotropic medium are the same as those shown in Fig. 1.7 for a wave in an isotropic medium. However, $\mathbf{E}_e \not\perp \mathbf{k}_e$ for an extraordinary wave in general, and $\mathbf{S}_e$ is not necessarily parallel to $\mathbf{k}_e$ because $\mathbf{D}_e \nparallel \mathbf{E}_e$. The only exception is when $\hat{e}_e$ is parallel to a principal axis. As a result, the direction of power flow, which is that of $\mathbf{S}_e$, is not the same as the direction of wavefront propagation, which is normal to the planes of constant phase and is that of $\mathbf{k}_e$. This is shown in Fig. 1.12(*b*) together with the relationships among the directions of the field vectors. Note that $\mathbf{E}_e$, $\mathbf{D}_e$, $\mathbf{k}_e$, and $\mathbf{S}_e$ lie in a plane normal to $\mathbf{H}_e$ because $\mathbf{B}_e \parallel \mathbf{H}_e$. Though (1.90) is still true according to (1.127), the relations between $\mathbf{E}$ and $\mathbf{H}$ in (1.92) are no longer valid for an extraordinary wave.



(*a*) (*b*)

Figure 1.12 Relationships among the directions of **E**, **D**, **H**, **B**, **k**, and **S** in an anisotropic medium for (*a*) an ordinary wave and (*b*) an extraordinary wave. In both cases, the vectors **E**, **D**, **k**, and **S** lie in a plane normal to **H**.

If the electric field of an extraordinary wave is not parallel to a principal axis, its Poynting vector is not parallel to its propagation direction because $\mathbf{E}_\mathrm{e}$ is not parallel to $\mathbf{D}_\mathrm{e}$. As a result, its energy flows away from the direction of its wavefront propagation. This phenomenon is known as *spatial beam walk-off*. If this characteristic appears in one of the two normal modes of an optical wave propagating in an anisotropic crystal, the optical wave will split into two beams of parallel wavevectors but separate, nonparallel traces of energy flow.

For simplicity, let us consider the propagation of an optical wave in a uniaxial crystal with $\hat{k}$, for both ordinary and extraordinary waves, at an angle θ with respect to the optical axis $\hat{z}$. Clearly, there is no walk-off for the ordinary wave because $\mathbf{E}_\mathrm{o} \parallel \mathbf{D}_\mathrm{o}$ and $\mathbf{S}_\mathrm{o} \parallel \hat{k}$. For the extraordinary wave, $\mathbf{S}_\mathrm{e}$ is not parallel to $\hat{k}$ but points in a direction at an angle ψ_e with respect to the optical axis. Figure 1.13(*a*) shows the relationships among these vectors. The angle α between $\mathbf{S}_\mathrm{e}$ and $\hat{k}$, which is defined as $\alpha = \psi_\mathrm{e} - \theta$, is called the *walk-off angle* of the extraordinary wave. Note that α is also the angle between $\mathbf{E}_\mathrm{e}$ and $\mathbf{D}_\mathrm{e}$, as can be seen from Fig. 1.13(*a*). Because neither $\mathbf{E}_\mathrm{e}$ nor $\mathbf{D}_\mathrm{e}$ is parallel to any principal axis, their relationship is found through their projections on the principal axes: $D_z^\mathrm{e} = n_\mathrm{e}^2 \epsilon_0 E_z^\mathrm{e}$ and $D_{xy}^\mathrm{e} = n_\mathrm{o}^2 \epsilon_0 E_{xy}^\mathrm{e}$. Using these two relations and the definition of α in Figs. 1.12(*b*) and 1.13(*a*), it can be shown that the walk-off angle is given by (see Problems 1.6.14 and 1.6.15)

$$\alpha = \psi_\mathrm{e} - \theta = \tan^{-1}\left(\frac{n_\mathrm{o}^2}{n_\mathrm{e}^2}\tan\theta\right) - \theta. \tag{1.131}$$

If the crystal is positive uniaxial, α as defined in Fig. 1.13(*a*) is negative. This means that $\mathbf{S}_\mathrm{e}$ is between $\hat{k}$ and $\hat{z}$ for a positive uniaxial crystal. If the crystal is negative uniaxial, α is positive and $\hat{k}$ is between $\mathbf{S}_\mathrm{e}$ and $\hat{z}$. No walk-off appears if an optical wave propagates along any of the principal axes of a crystal.

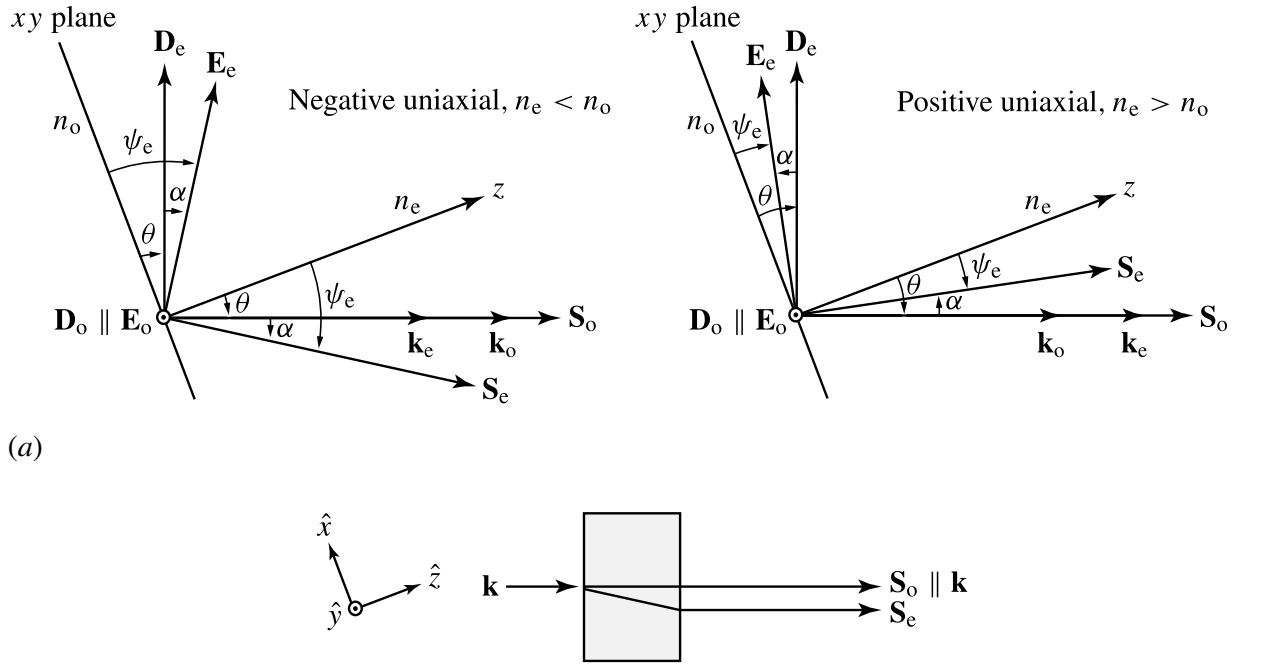


Figure 1.13 (*a*) Wave propagation and walk-off in a uniaxial crystal. (*b*) Birefringent plate acting as a polarizing beam splitter for a normally incident wave. The $\hat{x}$, $\hat{y}$, and $\hat{z}$ unit vectors indicate the principal axes of the birefringent plate.

A birefringent crystal can be used to construct a very simple *polarizing beam splitter* by taking advantage of the walk-off phenomenon. For such a purpose, a uniaxial crystal can be cut into a plate whose surfaces are at an oblique angle with respect to the optical axis, as is shown in Fig. 1.13(*b*). When an optical wave is normally incident upon the plate, it splits into ordinary and extraordinary waves in the crystal if its original polarization contains components of both polarizations. The extraordinary wave is then separated from the ordinary wave because of spatial walk-off, creating two orthogonally polarized beams. However, because of normal incidence, both $\mathbf{k}_e$ and $\mathbf{k}_o$ are parallel to the direction of $\hat{k}$ although they have different magnitudes. When both beams reach the other side of the plate, they are separated by a distance of $d = l \tan\alpha$, where l is the thickness of the plate. After leaving the plate, the two spatially separated beams propagate parallel to each other along the same direction $\hat{k}$ because the directions of their wavevectors have not changed, as is also shown in Fig. 1.13(*b*).

EXAMPLE 1.6 Find the spatial walk-off angle at 1 μm wavelength at a few representative propagation directions in KDP. Design a polarizing beam splitter at this wavelength using a KDP crystal.

Solution For a KDP crystal, $n_o = 1.51$ and $n_e = 1.47$ at 1 μm wavelength. The spatial walk-off angle α of an extraordinary wave is a function of the angle θ between the wave

propagation direction $\hat{k}$ and the optical axis $\hat{z}$ of the crystal. For example,

$$\alpha = \tan^{-1}\left(\frac{1.51^2}{1.47^2}\tan 30^\circ\right) - 30^\circ = 1.35^\circ, \qquad \text{for } \theta = 30^\circ,$$

$$\alpha = \tan^{-1}\left(\frac{1.51^2}{1.47^2}\tan 45^\circ\right) - 45^\circ = 1.54^\circ, \qquad \text{for } \theta = 45^\circ,$$

$$\alpha = \tan^{-1}\left(\frac{1.51^2}{1.47^2}\tan 60^\circ\right) - 60^\circ = 1.31^\circ, \qquad \text{for } \theta = 60^\circ.$$

From these numerical examples, we find that the walk-off angle does not vary monotonically with θ (see Problem 1.6.15).

A polarizing beam splitter can be made by cutting a KDP crystal at an angle, such as 45°, with respect to its optical axis for a parallel plate of thickness l. A beam at 1 μm wavelength that consists of a mix of extraordinary and ordinary polarizations is normally incident on the plate for $\theta = 45^\circ$ and $\alpha = 1.54^\circ$. Because the ordinary wave does not have walk-off, the Poynting vectors of the extraordinary and ordinary components of the beam separate at an angle of $\alpha = 1.54^\circ$. If a minimum spatial separation of $d = 100$ μm between the extraordinary and ordinary components is desired on the exit surface of the KDP plate, the minimum thickness of the plate has to be $l > d/\tan\alpha = 3.7$ mm.

Optical anisotropy and crystal symmetry

The optical anisotropy of a crystal depends on its structural symmetry. Crystals are classified into seven systems according to their symmetry. The linear optical properties of these seven systems are summarized in Table 1.2. Some important remarks regarding the relation between the optical properties and the structural symmetry of a crystal are made:

1. A cubic crystal need not have an isotropic structure although its linear optical properties are isotropic. For example, most III–V semiconductors, such as GaAs, InP, InAs, AlAs, etc., are cubic crystals with isotropic linear optical properties. Nevertheless, they have well-defined crystal axes, $\hat{a}$, $\hat{b}$, and $\hat{c}$. They are also polar semiconductors, which have anisotropic nonlinear optical properties.

Table 1.2 *Linear optical properties of crystals*

Crystal symmetry	Optical property
Cubic	Isotropic: $n_x = n_y = n_z$
Trigonal, tetragonal, hexagonal	Uniaxial: $n_x = n_y \neq n_z$
Orthorhombic, monoclinic, triclinic	Biaxial: $n_x \neq n_y \neq n_z$

2. Although the principal axes may coincide with the crystal axes in certain crystals, they are not the same concept. The crystal axes, denoted by $\hat{a}$, $\hat{b}$, and $\hat{c}$, are defined by the structural symmetry of a crystal, whereas the principal axes, denoted by $\hat{x}$, $\hat{y}$, and $\hat{z}$, are determined by the symmetry of $\boldsymbol{\epsilon}$. The principal axes of a crystal are orthogonal to one another, but the crystal axes are not necessarily so.

1.7 Gaussian beam

Because the wave equation governs optical propagation, the transverse field distribution pattern and its variation along the longitudinal propagation direction have to satisfy this equation in order for the wave to exist and to propagate. A well-defined field pattern that can remain unchanged as the wave propagates is called a *mode* of wave propagation. Such a transverse field pattern is known as a *transverse mode*. The optical modes that exist in a given medium are determined by the optical properties of the medium together with any boundary conditions imposed on the wave equation by the optical structures in the medium. Here we consider the optical modes in a homogeneous medium. Modes in waveguides and optical fibers are discussed in Chapters 2 and 3.

A monochromatic optical wave propagating in an isotropic, homogeneous medium is governed by the wave equation given in (1.82). Clearly, the monochromatic plane wave expressed in (1.58) is a solution of this wave equation. Therefore, plane waves are normal modes in an isotropic, homogeneous medium. They are not the only normal modes, however, as the wave equation governing wave propagation in such a medium has other normal-mode solutions. One such important set of modes is the *Gaussian modes*. Like plane waves, Gaussian modes are normal modes of wave propagation in an isotropic, homogeneous medium. Different from a plane wave, however, a Gaussian mode has a finite cross-sectional field distribution defined by its *spot size*. Being an unguided field with a finite spot size, a Gaussian mode differs from a waveguide mode, discussed in Chapters 2 and 3, in that its spot size varies along its longitudinal axis, taken to be the z axis, of propagation though its pattern remains unchanged. Therefore, its transverse field distribution also changes with z though the field pattern does not change. A Gaussian mode field at a frequency ω can thus be expressed as

$$\mathbf{E}_{mn}(\mathbf{r}, t) = \boldsymbol{\mathcal{E}}_{mn}(x, y, z)\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r} - \mathrm{i}\omega t) = \hat{e}\mathcal{E}_{mn}(x, y, z)\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r} - \mathrm{i}\omega t), \quad (1.132)$$

with a corresponding field distribution for its magnetic field component, where m and n are mode indices associated with the two transverse dimensions x and y, respectively. A Gaussian mode field has neither longitudinal electric nor longitudinal magnetic field components. It is a *TEM mode* that has only transverse electric and magnetic field components. Normal modes are orthonormal to each other and can be normalized, as

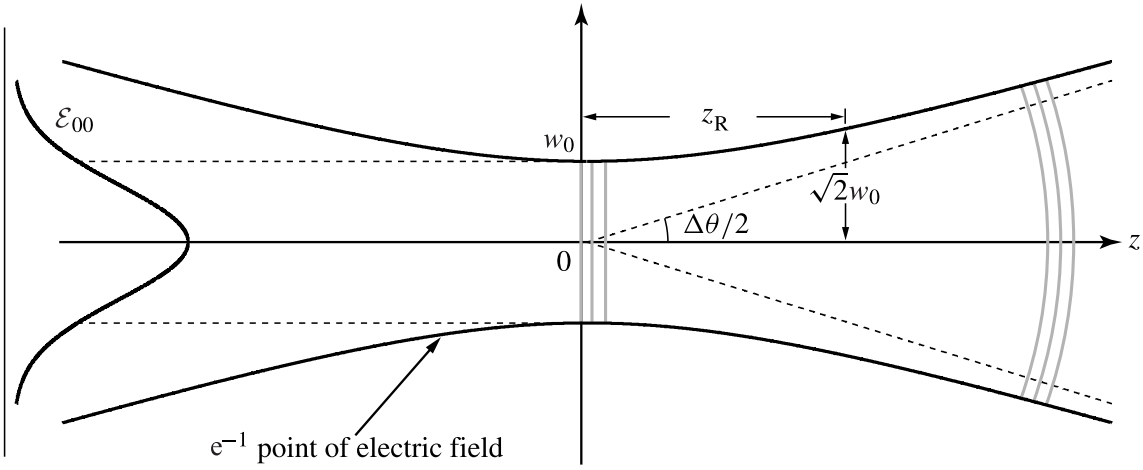


Figure 1.14 Gaussian beam characteristics.

is discussed in detail in Section 2.4. Gaussian modes are normalized by the following condition:

$$\frac{2k}{\omega\mu_0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\mathcal{E}_{mn}(x,y,z)|^2\mathrm{d}x\mathrm{d}y=1. \tag{1.133}$$

The location, taken to be $z = 0$ for a beam propagating along the z axis, where the smallest spot size of the beam occurs, is known as the *waist* of a Gaussian beam. The minimum *Gaussian beam spot size*, w_0, is defined as the e^{-2} radius of the Gaussian beam intensity profile at the beam waist. The diameter of the beam waist is $d_0 = 2w_0$. As illustrated in Fig. 1.14, a Gaussian beam has a plane wavefront at its beam waist. The beam remains well collimated within a distance of

$$z_R=\frac{kw_0^2}{2}=\frac{\pi n w_0^2}{\lambda}, \tag{1.134}$$

known as the *Rayleigh range*, on either side of the beam waist. In (1.134), $k = 2\pi n/\lambda$ is the propagation constant of the optical beam in a medium of refractive index n. The parameter $b = 2z_R$ is called the *confocal parameter* of the Gaussian beam. Because of diffraction, a Gaussian beam diverges away from its waist and acquires a spherical wavefront. As a result, both its spot size, $w(z)$, and the radius of curvature, $\mathcal{R}(z)$, of its wavefront are functions of distance z from its beam waist:

$$w(z)=w_0\left(1+\frac{z^2}{z_R^2}\right)^{1/2}=w_0\left[1+\left(\frac{2z}{kw_0^2}\right)^2\right]^{1/2} \tag{1.135}$$

and

$$\mathcal{R}(z)=z\left(1+\frac{z_R^2}{z^2}\right)=z\left[1+\left(\frac{kw_0^2}{2z}\right)^2\right]. \tag{1.136}$$

We see from (1.135) that at $z = \pm z_R$, $w = \sqrt{2}w_0$. At $|z| \gg z_R$, far away from the beam waist, $\mathcal{R}(z) \approx z$ and $w(z) \approx 2|z|/kw_0$. Therefore, the far-field beam *divergence angle* is

$$\Delta\theta = 2\frac{w(z)}{|z|} = \frac{4}{kw_0} = \frac{2\lambda}{\pi n w_0}. \tag{1.137}$$

For the far field at $|z| \gg z_R$, we find that the beam spot size $w(z)$ is inversely proportional to the beam waist spot size w_0 but is linearly proportional to the distance $|z|$ from the beam waist. This characteristic does not exist for the near field at $|z| \leq z_R$.

A complete set of Gaussian modes includes the fundamental TEM_{00} mode and high-order TEM_{mn} modes. The specific forms of the mode fields depend on the transverse coordinates of symmetry: the mode fields are described by a set of Hermite–Gaussian functions in rectangular coordinates, whereas they are described by the Laguerre–Gaussian functions in cylindrical coordinates. Because there is no structurally determined symmetry in free space, either set is equally valid. Usually the Hermite–Gaussian functions in the rectangular coordinates are used. In a transversely isotropic and homogeneous medium, a normalized TEM_{mn} Hermite–Gaussian mode field propagating along the z axis can be expressed as

$$\begin{aligned}\hat{\mathcal{E}}_{mn}(x, y, z) &= \frac{A_{mn}}{w(z)} H_m\left[\frac{\sqrt{2}x}{w(z)}\right] H_n\left[\frac{\sqrt{2}y}{w(z)}\right] \exp\left[\mathrm{i}\frac{k}{2}\frac{x^2+y^2}{q(z)}\right] \exp\left[\mathrm{i}\zeta_{mn}(z)\right] \\ &= \frac{A_{mn}}{w(z)} H_m\left[\frac{\sqrt{2}x}{w(z)}\right] H_n\left[\frac{\sqrt{2}y}{w(z)}\right] \exp\left[-\frac{x^2+y^2}{w^2(z)}\right] \exp\left[\mathrm{i}\frac{k}{2}\frac{x^2+y^2}{\mathcal{R}(z)}\right] \\ &\quad \times \exp\left[\mathrm{i}\zeta_{mn}(z)\right], \end{aligned} \tag{1.138}$$

where $A_{mn} = (\omega\mu_0/\pi k)^{1/2}(2^{m+n}\, m!\, n!)^{-1/2}$ is the normalization constant, H_m is the Hermite polynomial of order m, $q(z)$ is the complex radius of curvature of the Gaussian wave,

$$q(z) = z - \mathrm{i}z_R \quad \text{or} \quad \frac{1}{q(z)} = \frac{1}{\mathcal{R}(z)} + \mathrm{i}\frac{2}{kw^2(z)}, \tag{1.139}$$

and $\zeta_{mn}(z)$ is a mode-dependent on-axis phase variation along the z axis given by

$$\zeta_{mn}(z) = -(m+n+1)\tan^{-1}\frac{z}{z_R} = -(m+n+1)\tan^{-1}\left(\frac{2z}{kw_0^2}\right). \tag{1.140}$$

The Hermite polynomials can be obtained using the following relation:

$$H_m(\xi) = (-1)^m \mathrm{e}^{\xi^2}\frac{\mathrm{d}^m \mathrm{e}^{-\xi^2}}{\mathrm{d}\xi^m}. \tag{1.141}$$

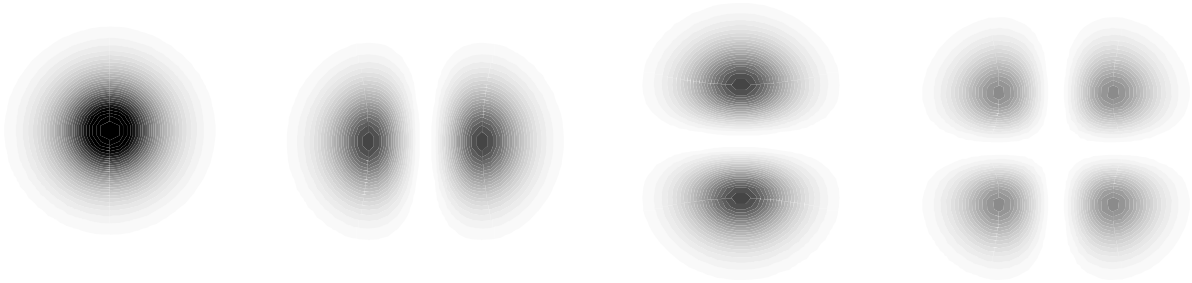


Figure 1.15 Intensity patterns of Hermite–Gaussian modes.

Some low-order Hermite polynomials are

$$H_0(\xi) = 1, \qquad H_1(\xi) = 2\xi, \qquad H_2(\xi) = 4\xi^2 - 2, \qquad H_3(\xi) = 8\xi^3 - 12\xi. \quad (1.142)$$

We see from (1.138) and (1.142) that the transverse field distribution $|\hat{\mathcal{E}}_{00}(x, y)|$ of the fundamental Gaussian mode, TEM_{00}, at any fixed longitudinal location z is simply a Gaussian function of the transverse radial distance $r = (x^2 + y^2)^{1/2}$ and that the spot size $w(z)$ is the e^{-1} radius of this Gaussian field distribution at z. The transverse field distribution of a high-order mode, TEM_{mn}, is the same Gaussian distribution spatially modulated by the Hermite polynomials H_m in x and H_n in y. As a result, its field distribution is more spread out radially than that of the fundamental TEM_{00} mode. In general, the higher the order of a mode, the farther its transverse field distribution spreads out. The intensity patterns of some Hermite–Gaussian modes are shown in Fig. 1.15.

EXAMPLE 1.7 A fundamental Gaussian beam in free space at the He–Ne laser wavelength of 632.8 nm has a spot size of $w_0 = 500$ μm at its beam waist. This beam has a Rayleigh range $z_R = \pi w_0^2/\lambda = 1.24$ m and a confocal parameter $b = 2z_R = 2.48$ m. Using (1.135) and (1.136), we find the following spot sizes and radii of curvature at a few different locations:

$w = 502$ μm, $\mathcal{R} = \pm 15.5$ m at $z = \pm 10$ cm,
$w = 642$ μm, $\mathcal{R} = \pm 2.54$ m at $z = \pm 1$ m,
$w \approx 40$ cm, $\mathcal{R} \approx \pm 1$ km at $z = \pm 1$ km.

From these numerical examples, we see that a Gaussian beam diverges very slowly, much like a plane wave, within the Rayleigh range on both sides of its beam waist. At the beam waist, a Gaussian beam has a plane wavefront with $\mathcal{R} = \infty$. At a distance much larger than the Rayleigh range on either side of the beam waist, a Gaussian beam approaches the characteristics of a spherical wave with $\mathcal{R} \approx z$. The Gaussian beam in this example has a far-field divergence angle of $\Delta\theta = 2\lambda/\pi w_0 = 0.8$ mrad.

1.8 Reflection and refraction

The characteristics of reflection and refraction of an optical wave at the interface of two different media depend on the properties of the media. We first consider the simple case of reflection and refraction at the planar interface of two dielectric media that are linear, lossless, and isotropic. In this situation, the permittivities ϵ_1 and ϵ_2 of the two media are constant real scalars, while the permeabilities are simply equal to μ_0 at optical frequencies. We assume that the optical wave is incident from medium 1 with a wavevector $\mathbf{k}_\mathrm{i}$, while the reflected wave has a wavevector $\mathbf{k}_\mathrm{r}$ and the transmitted wave has a wavevector $\mathbf{k}_\mathrm{t}$.

Because an optical wave varies with $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\mathrm{i}\omega t)$, the condition that

$$\mathbf{k}_\mathrm{i}\cdot\mathbf{r}=\mathbf{k}_\mathrm{r}\cdot\mathbf{r}=\mathbf{k}_\mathrm{t}\cdot\mathbf{r} \tag{1.143}$$

is required at the interface for the boundary conditions described by (1.17)–(1.20) to be satisfied at all points along the interface at all times. This implies that the three vectors $\mathbf{k}_\mathrm{i}$, $\mathbf{k}_\mathrm{r}$, and $\mathbf{k}_\mathrm{t}$ lie in the same plane known as the *plane of incidence*, as shown in Figs. 1.16 and 1.17. The projections of these three wavevectors on the interface are all equal so that

$$k_\mathrm{i}\sin\theta_\mathrm{i}=k_\mathrm{r}\sin\theta_\mathrm{r}=k_\mathrm{t}\sin\theta_\mathrm{t}, \tag{1.144}$$

where θ_i is the *angle of incidence*, and θ_r and θ_t are the angle of reflection and the angle of refraction, respectively, for the reflected and transmitted waves. All three angles are measured with respect to the normal $\hat{n}$ of the interface, as is shown in Figs. 1.16 and 1.17. Because $k_\mathrm{i}=k_\mathrm{r}$ and $k_\mathrm{i}/k_\mathrm{t}=n_1/n_2$, (1.144) yields the relation

$$\theta_\mathrm{i}=\theta_\mathrm{r} \tag{1.145}$$

and the following familiar *Snell law* for refraction:

$$n_1\sin\theta_\mathrm{i}=n_2\sin\theta_\mathrm{t}. \tag{1.146}$$

By expressing $\mathbf{H}$ in terms of $\mathbf{k}\times\mathbf{E}$ in the form of (1.86) with appropriate values of $\mathbf{k}$ for the incident, reflected, and refracted fields, the amplitudes of the reflected and transmitted fields can be obtained from the boundary conditions in (1.17) and (1.18). There are two different modes of field polarization.

TE polarization (s wave, σ wave)

The electric field is linearly polarized in a direction *perpendicular* to the plane of incidence while the magnetic field is polarized parallel to the plane of incidence, as

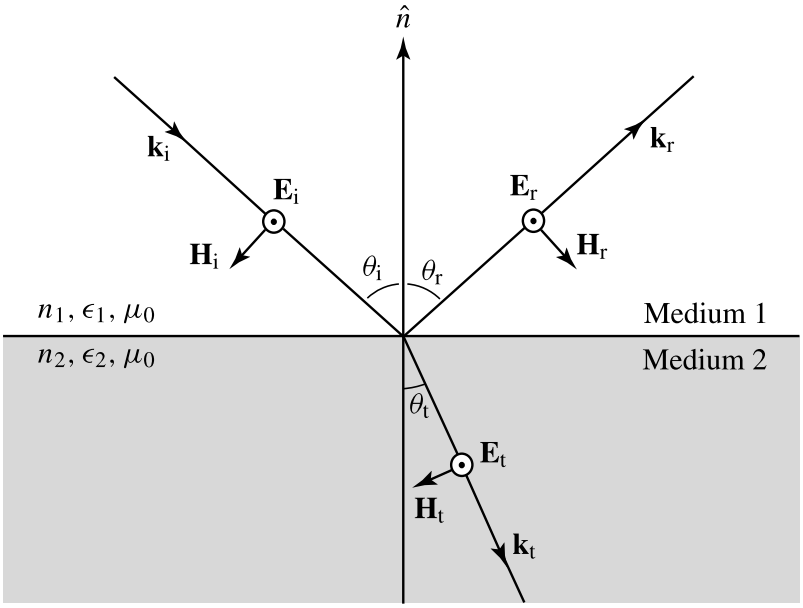


Figure 1.16 Reflection and refraction of a TE-polarized wave at the interface of two isotropic dielectric media. The three vectors $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{k}_t$ lie in the plane of incidence. The relationship between θ_i and θ_t shown here is for the case $n_1 < n_2$.

shown in Fig. 1.16. This is called *transverse electric (TE) polarization* or *perpendicular polarization*. This wave is also called *s polarized*, or σ *polarized*. In this case, the *reflection coefficient*, r, and the *transmission coefficient*, t, of the electric field are given by the following *Fresnel equations*:

$$r_s \equiv \frac{\mathcal{E}_r}{\mathcal{E}_i} = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} = \frac{n_1 \cos\theta_i - \sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}{n_1 \cos\theta_i + \sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}, \tag{1.147}$$

$$t_s \equiv \frac{\mathcal{E}_t}{\mathcal{E}_i} = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + \sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}, \tag{1.148}$$

respectively. The intensity *reflectance* and *transmittance*, R and T, which are also known as *reflectivity* and *transmissivity*, respectively, are given by

$$R_s \equiv \frac{I_r}{I_i} = \frac{|\overline{\mathbf{S}}_r \cdot \hat{n}|}{|\overline{\mathbf{S}}_i \cdot \hat{n}|} = \left| \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} \right|^2, \tag{1.149}$$

$$T_s \equiv \frac{I_t}{I_i} = \frac{|\overline{\mathbf{S}}_t \cdot \hat{n}|}{|\overline{\mathbf{S}}_i \cdot \hat{n}|} = 1 - R_s. \tag{1.150}$$

TM polarization (p wave, π wave)

The electric field is linearly polarized in a direction *parallel* to the plane of incidence while the magnetic field is polarized perpendicular to the plane of incidence, as shown

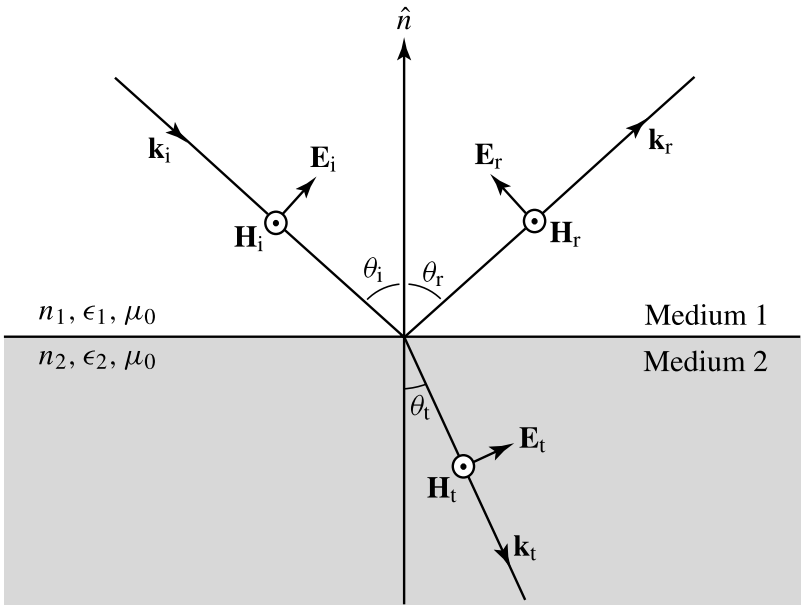


Figure 1.17 Reflection and refraction of a TM-polarized wave at the interface of two isotropic dielectric media. The three vectors $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{k}_t$ lie in the plane of incidence. The relationship between θ_i and θ_t shown here is for the case $n_1 < n_2$.

in Fig. 1.17. This is called *transverse magnetic (TM) polarization* or *parallel polarization*. This wave is also called *p polarized*, or π *polarized*. In this case, the reflection and transmission coefficients of the electric field are given by the following Fresnel equations:

$$r_p \equiv \frac{\mathcal{E}_r}{\mathcal{E}_i} = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} = \frac{n_2^2 \cos\theta_i - n_1\sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}{n_2^2 \cos\theta_i + n_1\sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}, \tag{1.151}$$

$$t_p \equiv \frac{\mathcal{E}_t}{\mathcal{E}_i} = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} = \frac{2n_1 n_2 \cos\theta_i}{n_2^2 \cos\theta_i + n_1\sqrt{n_2^2 - n_1^2 \sin^2\theta_i}}, \tag{1.152}$$

respectively. The intensity reflectance and transmittance for TM polarization are given, respectively, by

$$R_p \equiv \frac{I_r}{I_i} = \left|\frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t}\right|^2, \tag{1.153}$$

$$T_p \equiv \frac{I_t}{I_i} = 1 - R_p. \tag{1.154}$$

Several important characteristics of the reflection and refraction of an optical wave at an interface between two media are summarized.

1. For both TE and TM polarizations, $R = |r|^2$ and $R + T = 1$, but $T \neq |t|^2$.
2. If $n_1 < n_2$, light is incident from a rare medium upon a dense medium. In this case, the reflection is called *external reflection*. If $n_1 > n_2$, light is incident from a dense medium on a rare medium, and the reflection is called *internal reflection*.
3. **Normal incidence.** In the case of normal incidence, $\theta_i = \theta_t = 0$. There is no difference between TE and TM polarizations, and

$$R = \left|\frac{n_1 - n_2}{n_1 + n_2}\right|^2, \quad T = 1 - R = \frac{4n_1 n_2}{(n_1 + n_2)^2}. \tag{1.155}$$

   For the case of external reflection at normal incidence, there is a 180° phase reversal for the reflected electric field with respect to the incident field. For internal reflection, the phase of the reflected field is not reversed at normal incidence. However, the values of R and T do not depend on which side of the interface the wave comes from.
4. **Brewster angle.** For a TE wave, R_s increases monotonically with the angle of incidence. For a TM wave, R_p first decreases then increases as the angle of incidence increases. For the interface between two lossless media, $R_p = 0$ at an angle of incidence $\theta_i = \theta_B$, where

$$\theta_B = \tan^{-1}\frac{n_2}{n_1} \tag{1.156}$$

   is known as the *Brewster angle*. When $\theta_i = \theta_B$, the angle of refraction for the transmitted wave is

$$\theta_t = \frac{\pi}{2} - \theta_B. \tag{1.157}$$

   It can be shown that this angle is the Brewster angle for the same wave incident from the other side of the interface. Figure 1.18(a) shows the reflectance of TE and TM waves as a function of the angle of incidence for external reflection at the interface between two media of refractive indices of 1 and 3.5. These characteristics are very useful in practical device applications: (a) at $\theta_i = \theta_B$, a TM-polarized wave is totally transmitted, resulting in a perfect lossless window for TM polarization – such windows are called *Brewster windows* and are useful as laser windows; (b) at $\theta_i = \theta_B$, the reflected wave is completely TE polarized – linearly polarized light can be produced by a *reflection-type polarizer* based on this principle.
5. **Critical angle.** In the case of internal reflection with $n_1 > n_2$, *total internal reflection* occurs if the angle of incidence θ_i is larger than the angle

$$\theta_c = \sin^{-1}\frac{n_2}{n_1}, \tag{1.158}$$

   which is called the *critical angle*. The reflectance of TE and TM waves as a function of angle of incidence for internal reflection at the interface between two media of

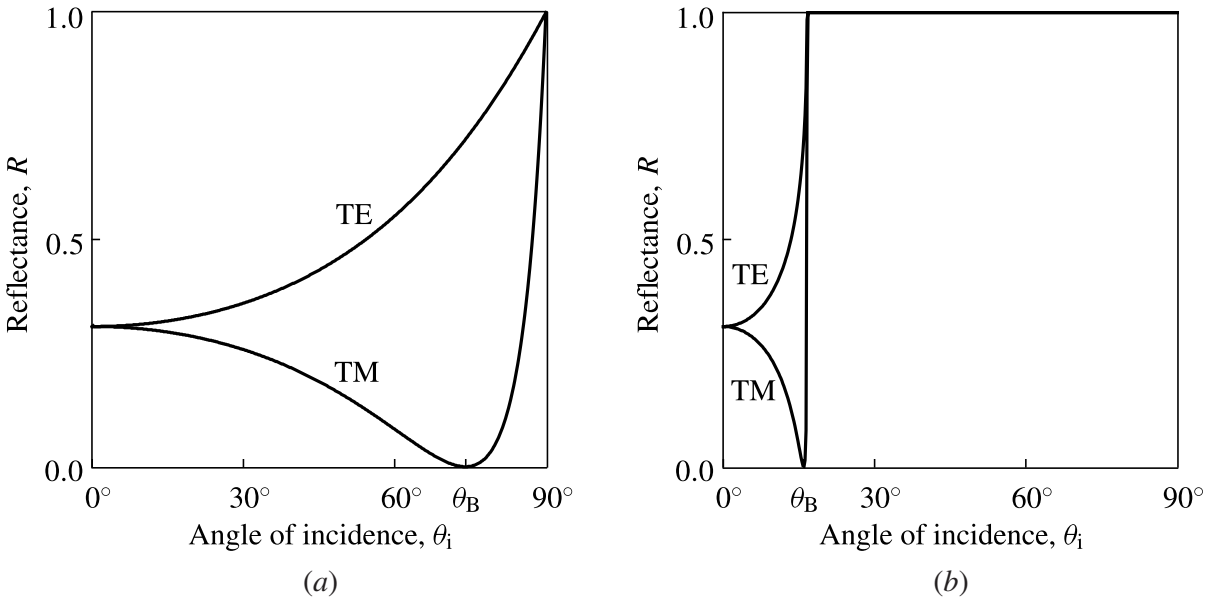


Figure 1.18 Reflectance of TE and TM waves at an interface of lossless media as a function of the angle of incidence for (*a*) external reflection and (*b*) internal reflection.

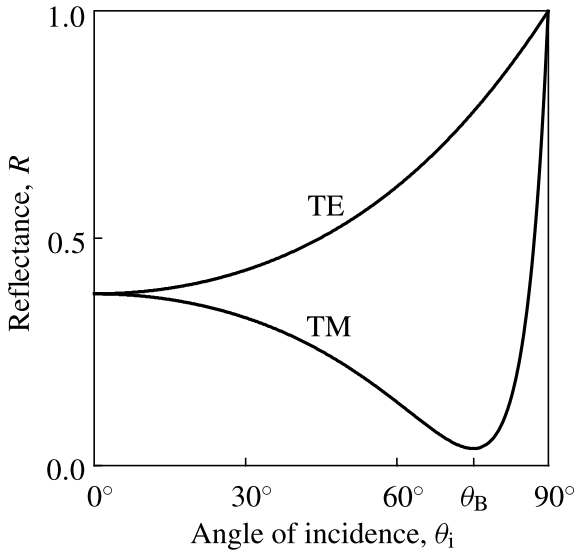


Figure 1.19 Reflectance of TE and TM waves at an interface of lossy or amplifying media as a function of the angle of incidence for external reflection.

refractive indices of 1 and 3.5 is shown in Fig. 1.18(*b*). Note that the Brewster angle for internal reflection is always smaller than the critical angle.

6. If one or both media have a loss or gain, the indices of refraction become complex. In this situation, the reflectance of the TM wave has a minimum that does not reach zero, as shown in Fig. 1.19 for external reflection.
7. For wave propagation in a general direction in an anisotropic medium, there are two normal modes that have different indices of refraction. The refracted fields of these two normal modes can propagate in different directions, resulting in the

phenomenon of *double refraction*. Meanwhile, the Poynting vector does not have to be in the plane of incidence.

8. Optical media are generally dispersive. Therefore, reflectance and transmittance, as well as the direction of the refracted wave, are generally frequency dependent.

EXAMPLE 1.8 Water has an index of refraction $n = 1.33$. The index of refraction of ordinary glass is approximately $n = 1.5$. For most semiconductors, such as Si, GaAs, and InP, the index of refraction is often in the range between 3 and 4, depending on the optical wavelength and the material. Here we take a nominal value of $n = 3.5$ for a semiconductor. Find the reflectivities at normal incidence, the Brewster angles, and the critical angles for these media at their interfaces with air.

Solution Using the formula given in (1.155) for the reflectivity at normal incidence, we find that $R = 0.02$ for water, $R = 0.04$ for ordinary glass, and R typically falls in the range of 0.3 and 0.32 for a semiconductor. Using (1.156) for the Brewster angle, we find that $\theta_B \approx 54^\circ$ for water, $\theta_B \approx 56^\circ$ for ordinary glass, and θ_B is typically around 74° for a semiconductor. Using (1.158) for the critical angle, we find that $\theta_c \approx 49^\circ$ for water, $\theta_c \approx 42^\circ$ for ordinary glass, and θ_c is around 17° for a semiconductor.

1.9 Phase velocity, group velocity, and dispersion

For a monochromatic plane optical wave traveling in the z direction, the electric field can be written as

$$\mathbf{E} = \boldsymbol{\mathcal{E}} \exp(\mathrm{i}kz - \mathrm{i}\omega t), \tag{1.159}$$

where $\boldsymbol{\mathcal{E}}$ is a constant vector independent of space and time. This represents a sinusoidal wave whose phase varies with z and t as

$$\varphi = kz - \omega t. \tag{1.160}$$

For a point of constant phase on the space- and time-varying field, $\varphi =$ constant and thus $k\mathrm{d}z - \omega \mathrm{d}t = 0$. If we track this point of constant phase, we find that it is moving with a velocity of

$$v_p = \frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\omega}{k}. \tag{1.161}$$

This is called the *phase velocity* of the wave. Note that the phase velocity is a function of optical frequency because the refractive index of a medium is a function of frequency. There is *phase-velocity dispersion* due to the fact that $\mathrm{d}n/\mathrm{d}\omega \neq 0$. In the case of *normal dispersion*, $\mathrm{d}n/\mathrm{d}\omega > 0$ and $\mathrm{d}n/\mathrm{d}\lambda < 0$; in the case of *anomalous dispersion*, $\mathrm{d}n/\mathrm{d}\omega < 0$ and $\mathrm{d}n/\mathrm{d}\lambda > 0$.

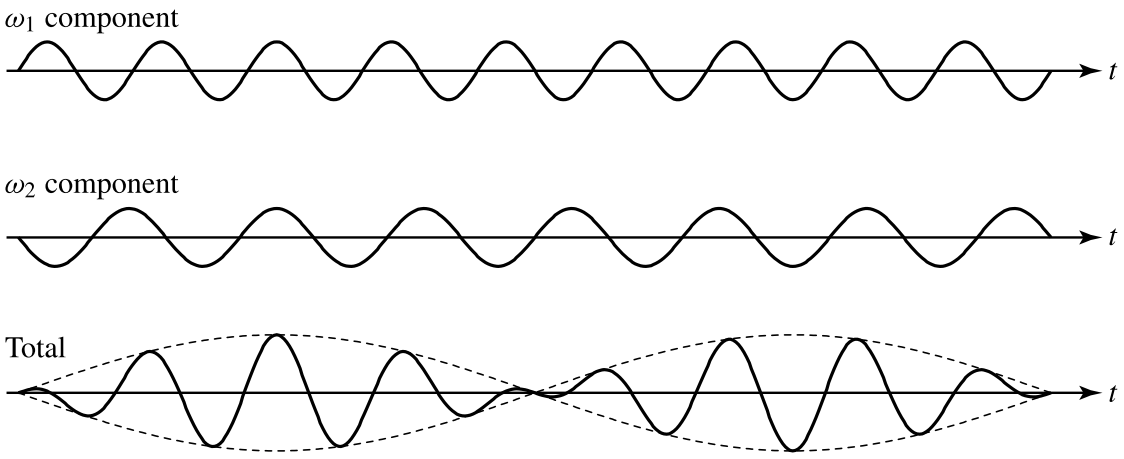


Figure 1.20 Wave packet composed of two frequency components showing the carrier and the envelope. The carrier travels at the phase velocity, whereas the envelope travels at the group velocity.

In real circumstances, a propagating optical wave rarely contains only one frequency. It usually consists of many frequency components grouped around some center frequency ω_0. For the simplicity of illustration, we consider a wave packet traveling in the z direction that is composed of two plane waves of equal real amplitude $\mathcal{E}$. The frequencies and propagation constants of the two component plane waves are

$$\begin{aligned} &\omega_1 = \omega_0 + \mathrm{d}\omega,\ k_1 = k_0 + \mathrm{d}k, \\ &\omega_2 = \omega_0 - \mathrm{d}\omega,\ k_2 = k_0 - \mathrm{d}k. \end{aligned} \tag{1.162}$$

The space- and time-dependent total real field of the wave packet is then given by

$$\begin{aligned} \boldsymbol{E} &= \mathcal{E}\exp(\mathrm{i}k_1 z - \mathrm{i}\omega_1 t) + \text{c.c.} + \mathcal{E}\exp(\mathrm{i}k_2 z - \mathrm{i}\omega_2 t) + \text{c.c.} \\ &= 2\mathcal{E}\{\cos[(k_0 + \mathrm{d}k)z - (\omega_0 + \mathrm{d}\omega)t] + \cos[(k_0 - \mathrm{d}k)z - (\omega_0 - \mathrm{d}\omega)t]\} \\ &= 4\mathcal{E}\cos(\mathrm{d}kz - \mathrm{d}\omega t)\cos(k_0 z - \omega_0 t). \end{aligned} \tag{1.163}$$

We find that the resultant wave packet has a *carrier*, which has a frequency ω_0 and a propagation constant k_0, and an *envelope*, which varies in space and time as $\cos(\mathrm{d}kz - \mathrm{d}\omega t)$. This is illustrated in Fig. 1.20. Therefore, a fixed point on the envelope is defined by $\mathrm{d}kz - \mathrm{d}\omega t = \text{constant}$, and it travels with a velocity of

$$v_\mathrm{g} = \frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\mathrm{d}\omega}{\mathrm{d}k}. \tag{1.164}$$

This is the velocity of the wave packet and is called the *group velocity*. Because the energy of a harmonic wave is proportional to the square of its field amplitude, the energy carried by a wave packet that is composed of many frequency components is concentrated in regions where the amplitude of the envelope is large. Therefore, the energy in a wave packet is transported at group velocity v_g. *The constant-phase wavefront travels at the phase velocity, but the group velocity is the velocity at which energy and information travel.*

In reality, group velocity is usually a function of optical frequency. Then,

$$\frac{\mathrm{d}^2k}{\mathrm{d}\omega^2} = \frac{\mathrm{d}}{\mathrm{d}\omega}v_{\mathrm{g}}^{-1} \neq 0. \tag{1.165}$$

Therefore, $\mathrm{d}^2k/\mathrm{d}\omega^2$ represents *group-velocity dispersion*. A *dimensionless* coefficient for group-velocity dispersion can be defined as

$$D = c\omega\frac{\mathrm{d}^2k}{\mathrm{d}\omega^2} = \frac{2\pi c^2}{\lambda}\frac{\mathrm{d}^2k}{\mathrm{d}\omega^2}. \tag{1.166}$$

Group-velocity dispersion is an important consideration in the propagation of optical pulses. It can cause broadening of an individual pulse, as well as changes in the time delay between pulses of different frequencies. The sign of the group-velocity dispersion can be either positive or negative. In the case of *positive group-velocity dispersion*, $\mathrm{d}^2k/\mathrm{d}\omega^2 > 0$ and $D > 0$, a long-wavelength, or low-frequency, pulse travels faster than a short-wavelength, or high-frequency, pulse. In contrast, a short-wavelength pulse travels faster than a long-wavelength pulse in the case of *negative group-velocity dispersion*, $\mathrm{d}^2k/\mathrm{d}\omega^2 < 0$ and $D < 0$. In a given material, the sign of D generally depends on the spectral region of concern. Group-velocity dispersion and phase-velocity dispersion discussed earlier have different meanings. They should not be confused with each other.

When measuring the transmission delay or the broadening of optical pulses due to dispersion in optical fibers, another dispersion coefficient defined as

$$D_\lambda = -\frac{2\pi c}{\lambda^2}\frac{\mathrm{d}^2k}{\mathrm{d}\omega^2} = -\frac{D}{c\lambda} \tag{1.167}$$

is usually used. This coefficient is generally expressed as a function of wavelength in units of picoseconds per kilometer per nanometer. It is a direct measure of the chromatic pulse transmission delay over a unit transmission length.

In general, both $\epsilon(\omega)$ and $n(\omega)$ in an optical medium are frequency dependent, and the propagation constant is

$$k = \frac{\omega}{c}n(\omega). \tag{1.168}$$

Therefore, we have

$$v_{\mathrm{p}} = \frac{c}{n} \tag{1.169}$$

and

$$v_{\mathrm{g}} = \frac{c}{N}, \tag{1.170}$$

where

$$N = n + \omega\frac{\mathrm{d}n}{\mathrm{d}\omega} = n - \lambda\frac{\mathrm{d}n}{\mathrm{d}\lambda} \tag{1.171}$$

is called the *group index*. Using (1.166) and (1.167), we also have

$$D(\lambda) = \lambda^2 \frac{\mathrm{d}^2 n}{\mathrm{d}\lambda^2} \tag{1.172}$$

and

$$D_\lambda(\lambda) = -\frac{\lambda}{c} \frac{\mathrm{d}^2 n}{\mathrm{d}\lambda^2}, \tag{1.173}$$

respectively.

EXAMPLE 1.9 The index of refraction of a certain type of glass as a function of optical wavelength around $\lambda = 1.3$ μm can be approximated as $n = 1.465 - 0.0114(\lambda - 1.3) - 0.004(\lambda - 1.3)^3$, where λ is measured in micrometers. Therefore,

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} = -0.0114 - 0.012(\lambda - 1.3)^2,$$

$$N = n - \lambda \frac{\mathrm{d}n}{\mathrm{d}\lambda} = 1.48 + 0.0156(\lambda - 1.3)^2 + 0.008(\lambda - 1.3)^3,$$

$$D = \lambda^2 \frac{\mathrm{d}n^2}{\mathrm{d}\lambda^2} = -0.024\lambda^2(\lambda - 1.3).$$

We find that, in this spectral region, $\mathrm{d}n/\mathrm{d}\lambda < 0$ for any wavelength but $D > 0$ for $\lambda < 1.3$ μm and $D < 0$ for $\lambda > 1.3$ μm. Clearly, this glass has normal phase-velocity dispersion in the entire spectral region around $\lambda = 1.3$ μm, but it has positive group-velocity dispersion for $\lambda < 1.3$ μm and negative group-velocity dispersion for $\lambda > 1.3$ μm. As an example, we find that $n \approx 1.469$, $N \approx 1.481$, and $D \approx 0.0072$ at $\lambda = 1$ μm. We also find that $n \approx 1.463$, $N \approx 1.481$, and $D \approx -0.0108$ at $\lambda = 1.5$ μm. Because of normal phase-velocity dispersion, the group index is always larger than the refractive index, $N > n$, in this spectral region.

1.10 Material dispersion

As discussed in Sections 1.1 and 1.3, dispersion in the susceptibility of a medium is caused by the fact that the response of the medium to excitation by an optical field does not decay instantaneously. The general characteristics of the medium can be understood from its impulse response. In general, the impulse response of a medium decays exponentially while oscillating at some resonance frequencies. There may exist several exponential relaxation constants and several oscillation frequencies for a given material across the electromagnetic spectrum. This is true even within the optical spectral region. However, at a given optical frequency ω, the characteristics of the material response are dominated by the resonance frequency closest to ω and the relaxation constant associated with the oscillation at this particular resonance frequency. We therefore consider, for simplicity, a medium of a single resonance frequency at ω_0 with a relaxation

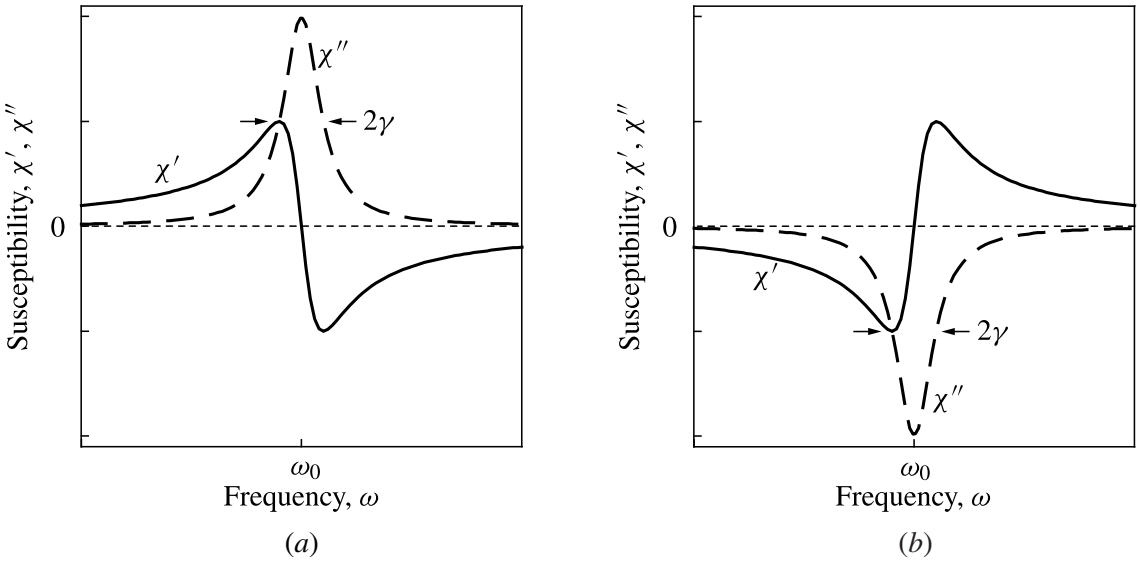


Figure 1.21 Real and imaginary parts, χ' and χ'', respectively, of susceptibility for a medium with (a) a loss and (b) a gain near a resonance frequency, ω_0.

constant γ. The susceptibility in the time domain is simply the impulse response of the medium, which is real and has the following general form:

$$\chi(t) \propto \begin{cases} \mathrm{e}^{-\gamma t} \sin \omega_0 t, & t > 0, \\ 0, & t < 0. \end{cases} \tag{1.174}$$

Note that $\chi(t) = 0$ for $t < 0$ because a medium can respond only after, but not before, an excitation. This is the *causality* condition, which applies to all physical systems.

The Fourier transform of (1.174) yields

$$\chi(\omega) = \int_{-\infty}^{\infty} \chi(t) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}t \approx -\chi_\mathrm{b} \frac{\omega_0}{\omega - \omega_0 + \mathrm{i}\gamma} \tag{1.175}$$

in the frequency domain, where $\chi_\mathrm{b} = \chi(\omega \ll \omega_0)$ is a constant equal to the background value of $\chi(\omega)$ at low frequencies far away from resonance. In (1.175), we have taken the so-called *rotating-wave approximation* by dropping a term that contains $\omega + \omega_0$ in its denominator because $\omega + \omega_0 \gg |\omega - \omega_0|$ in the optical spectral region (see Problem 1.10.1). This susceptibility has the following real and imaginary parts:

$$\chi'(\omega) = -\chi_\mathrm{b} \frac{\omega_0(\omega - \omega_0)}{(\omega - \omega_0)^2 + \gamma^2}, \qquad \chi''(\omega) = \chi_\mathrm{b} \frac{\omega_0 \gamma}{(\omega - \omega_0)^2 + \gamma^2}, \tag{1.176}$$

which are plotted in Fig. 1.21. Note that $\chi''(\omega)$ has a *Lorentzian lineshape*, which has a FWHM $\Delta\omega = 2\gamma$. The sign of χ'' depends on that of χ_b. In the normal state, $\chi_\mathrm{b} > 0$, and the medium has an optical loss near resonance. This characteristic results in the

absorption of light at frequency $\omega = \omega_0$. When $\chi_b < 0$, the medium has optical gain, resulting in the amplification of light at $\omega = \omega_0$ such as in the case of a laser. Note that both χ' and χ'' are proportional to χ_b. Therefore, *when χ'' changes sign, χ' also changes sign.* When $\chi'' < 0$, χ' is negative for $\omega < \omega_0$ and positive for $\omega > \omega_0$, as is shown in Fig. 1.21(*b*).

EXAMPLE 1.10 For an atomic transition associated with absorption or emission of optical radiation at 1 μm wavelength, the resonance frequency is $\nu_0 = c/\lambda = 300$ THz, thus $\omega_0 = 2\pi\nu_0 = 1.885 \times 10^{15}$ s^{-1}. If the polarization associated with this resonant transition relaxes with a time constant of $\tau = 1$ ps, then $\gamma = 1/\tau = 10^{12}$ s^{-1} and $\Delta\omega = 2\gamma = 2 \times 10^{12}$ s^{-1}. Thus the Lorentzian spectral line has a FWHM linewidth of $\Delta\nu = \Delta\omega/2\pi \approx 318$ GHz, which is considered quite broad but is approximately only 0.1% of the center frequency ν_0 of the spectral line. If the relaxation time constant is $\tau = 1$ ns, we find a spectral linewidth of $\Delta\nu \approx 318$ MHz. For a relaxation time constant of $\tau = 1$ μs, we have a narrow linewidth of $\Delta\nu \approx 318$ kHz.

Note that the spectral linewidth is determined by the *polarization relaxation time* rather than by the *population relaxation time* of a material. The polarization relaxation time constant is generally much smaller than the population relaxation time constant for a given transition. Therefore, the spectral linewidth of a given transition can be quite broad even when the energy levels involved have long population relaxation times. One good example is the optical transitions in Nd : YAG discussed in Section 10.1.

A medium generally has many resonance frequencies, each corresponding to an absorption frequency in the normal state of the medium. Because $\epsilon(\omega) = \epsilon_0(1 + \chi(\omega))$, the dispersion characteristics of $\epsilon(\omega)$ depend directly on those of $\chi(\omega)$ given by (1.176). Its real and imaginary parts in the normal state as a function of ω over a spectral range covering a few resonances are shown in Fig. 1.22. Some important dispersion characteristics of $\chi(\omega)$ and $\epsilon(\omega)$ are summarized below.

1. It can be seen from Fig. 1.21(*a*) that $\chi'(\omega \ll \omega_0)$ is larger than $\chi'(\omega \gg \omega_0)$ in the normal state. Therefore, around any single resonance frequency, ϵ' at any frequency on the low-frequency side has a value larger than that at any frequency on the high-frequency side.
2. A medium is said to have *normal dispersion* in a spectral region where ϵ' increases with frequency so that $d\epsilon'/d\omega > 0$. It is said to have *anomalous dispersion* in a spectral region where ϵ' decreases with increasing frequency so that $d\epsilon'/d\omega < 0$. Because $dn/d\omega$ and $d\epsilon'/d\omega$ have the same sign, the index of refraction also increases with frequency in a spectral region of normal dispersion and decreases with frequency in a spectral region of anomalous dispersion.

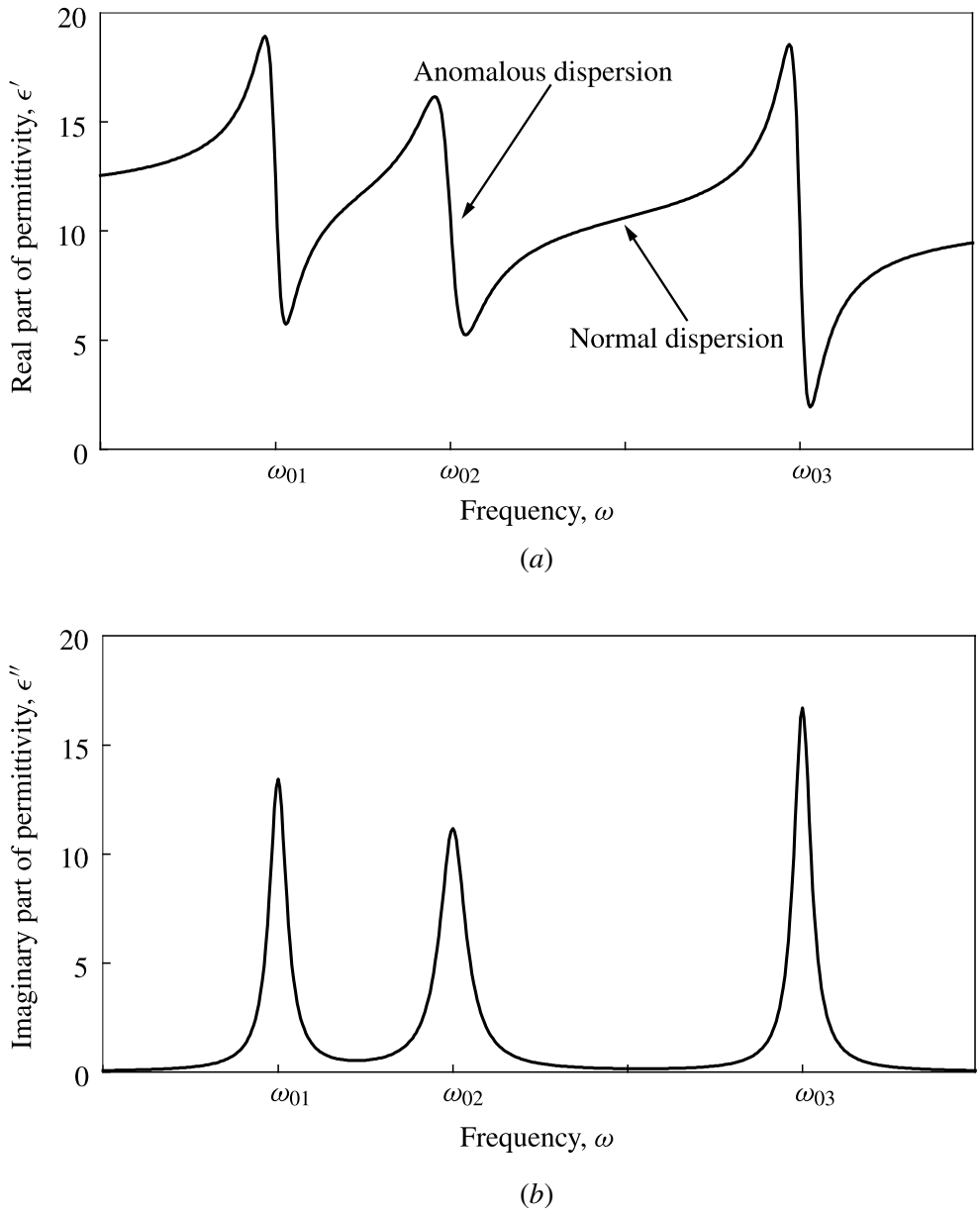


Figure 1.22 Real and imaginary parts of ϵ as a function of ω for a medium in its normal state over a spectral range covering a few resonance frequencies.

3. It can be seen from Fig. 1.22 that when a material is in its normal state, normal dispersion appears everywhere except in the immediate neighborhood within the FWHM of a resonance frequency where anomalous dispersion occurs. This characteristic can be reversed near a resonance frequency where resonant amplification, rather than absorption, exists.
4. Note the distinction between the definition of normal and anomalous dispersion in terms of the sign of $d\epsilon'/d\omega$ or $dn/d\omega$ and that of positive and negative group-velocity dispersion in terms of the sign of D. Both positive and negative

group-velocity dispersion can appear in a spectral region where the dispersion defined in terms of $\mathrm{d}n/\mathrm{d}\omega$ is normal.[3]

5. In most transparent materials, such as glass and water, normal dispersion appears in the visible spectral region and may extend to the near infrared and near ultraviolet regions.

Kramers–Kronig relations

It can be seen from the discussions above that the real and imaginary parts of $\chi(\omega)$, or those of $\epsilon(\omega)$, are not independent of each other. The susceptibility of any physical system has to satisfy the causality requirement in the time domain. This requirement leads to a general relationship between χ' and χ'' in the frequency domain:

$$\chi'(\omega) = \frac{2}{\pi}\mathrm{P}\int_0^\infty \frac{\omega'\chi''(\omega')}{\omega'^2-\omega^2}\mathrm{d}\omega', \qquad \chi''(\omega) = -\frac{2}{\pi}\mathrm{P}\int_0^\infty \frac{\omega\chi'(\omega')}{\omega'^2-\omega^2}\mathrm{d}\omega', \tag{1.177}$$

where the principal values are taken for the integrals. These relations are known as the *Kramers–Kronig relations*. Therefore, once the real part of $\chi(\omega)$ is known over the entire spectrum, its imaginary part can be found, and vice versa. Note that the relations in (1.177) are consistent with the fact that $\chi'(\omega)$ is an even function, while $\chi''(\omega)$ is an odd function, of ω, as discussed in Section 1.3. The contradiction to this statement seen in (1.176) is only apparent but not real. It is caused by the rotating-wave approximation taken in (1.175). There is no contradiction when the approximation is removed and exact expressions are used for $\chi'(\omega)$ and $\chi''(\omega)$ (see Problem 1.10.1).

1.11 Photon nature of light

When considering the function of a device that involves the emission or absorption of light, a purely electromagnetic wave description of light is not adequate. In this situation, the photon nature of light cannot be ignored. Meanwhile, the material involved in this process also undergoes quantum mechanical transitions between its energy levels.

The energy of a photon is determined by its frequency ν or, equivalently, its angular frequency ω. Associated with the particle nature of a photon, there is a momentum

[3] In the literature, positive group-velocity dispersion is sometimes referred to as normal dispersion while negative group-velocity dispersion is referred to as anomalous dispersion. This is confusing and is, strictly speaking, not correct.

determined by its wavelength λ or, equivalently, its wavevector $\mathbf{k}$. These characteristics are summarized below for a photon in free space:

$$\begin{aligned} &\text{speed} && c = \lambda\nu, \\ &\text{energy} && h\nu = \hbar\omega = pc, \\ &\text{momentum} && p = \frac{h\nu}{c} = \frac{h}{\lambda}, \qquad \mathbf{p} = \hbar\mathbf{k}. \end{aligned}$$

The energy of a photon that has a wavelength λ in free space can be calculated using the following formula:

$$h\nu = \frac{1.2398}{\lambda}\ \mu\text{m eV} = \frac{1239.8}{\lambda}\ \text{nm eV}. \tag{1.178}$$

For example, at an optical wavelength of 1 μm, the photon energy is 1.2398 eV.

The energy of a photon is determined only by the frequency, or wavelength, of light, but not by its intensity. The intensity of light is related to the flux density, or number per unit time per unit area, of photons by

$$\text{photon flux density} = \frac{I}{h\nu} = \frac{I}{\hbar\omega}. \tag{1.179}$$

EXAMPLE 1.11 It is found that a piece of crystal transmits light at $\lambda = 500$ nm but absorbs light at $\lambda = 400$ nm. Make an intelligent guess of its bandgap from this limited information.

Solution Because a crystal transmits photons with energies below its bandgap but absorbs those with energies above its bandgap, we can reasonably guess that the bandgap of this crystal falls between the photon energies corresponding to 500 and 400 nm wavelengths. Using (1.178) for the photon energy, we find that

$$2.48\ \text{eV} < E_g < 3.10\ \text{eV}.$$

PROBLEMS

1.1.1 Verify that Maxwell's equations and the continuity equation are invariant under (a) space inversion, (b) time reversal, and (c) space inversion and time reversal simultaneously.

1.3.1 Verify the reality condition for electric susceptibility and electric permittivity given in (1.56) and (1.57), respectively.

1.4.1 Two polarizers placed in tandem along the line of propagation of an optical beam are called *cross polarizers* if their axes are arranged to be orthogonal to each

other. For the purpose of answering the following questions, consider polarizers of transmission type.

a. Show that no light of any polarization can pass through a set of cross polarizers.
b. A third polarizer is inserted in between the two cross polarizers. The transmission of this three-polarizer combination is not zero any more if the axis of the inserted polarizer is not parallel to either of the original two. Find the transmittance of this combination as a function of the angle between the axis of this polarizer and that of the polarizer at the input end.
c. Since each polarizer acts only as a polarization-sensitive filter to transmit the field component of a particular polarization, the phenomenon described in (b) may not seem possible. Can you give a physically intuitive explanation for it?

1.5.1 Express the wavenumber β and the attenuation coefficient α defined in (1.100) for propagation of an optical wave in an absorptive medium in terms of the real part, χ', and the imaginary part, χ'', of the electric susceptibility of the medium. Show that when $\chi'' \ll \chi'$, we have

$$\alpha \approx \beta \frac{\chi''}{n^2}. \tag{1.180}$$

1.5.2 The electric susceptibility of pure crystalline silicon at the optical wavelength of $\lambda = 532$ nm is $\chi = 15.48 + \mathrm{i}0.284$. An optical beam of 1 W power at 532 nm wavelength is normally incident from the air on the surface of a crystalline silicon wafer, which is polished to mirror finish. The surface on the other side of the silicon wafer is antireflection coated so that no reflection of light takes place on that surface.

a. How much light (in milliwatts) is reflected from the surface from which the light enters the silicon wafer? How much enters the silicon wafer?
b. How much of the light entering the wafer is transmitted from the other side if the thickness of the silicon wafer is 100 μm?
c. What is the thickness of the wafer if 1 mW of light is transmitted from the other side?

1.6.1 An optical isolator transmits light traveling in one direction and blocks its reflection traveling in the opposite direction. Show that isolation of light reflected from a plane mirror can be accomplished by using a combination of a polarizer and a quarter-wave plate with the axis of the quarter-wave plate set at 45° with respect to the transmission axis of the polarizer.

1.6.2 A polarizer and a half-wave plate can be used to make an attenuator of linearly polarized light. Sketch a diagram of how this can be achieved and then plot the output intensity of the system as a function of the angle between the axis of the wave plate and that of the polarizer.

1.6.3 A crystal has the following electric permittivity tensor in the (x, y, z) coordinate system:

$$\epsilon = \epsilon_0 \begin{bmatrix} 2.25 & 0 & 0 \\ 0 & 2.13 & 0 \\ 0 & 0 & 2.02 \end{bmatrix}.$$

A linearly polarized optical wave that has a free-space wavelength $\lambda = 600$ nm is sent into the crystal. Find the wavelength of the wave in the crystal in each of the following arrangements.

a. The wave is polarized along $\hat{x}$ and propagates along $\hat{z}$.
b. The wave is polarized along $\hat{y}$ and propagates along $\hat{z}$.
c. The wave is polarized along $\hat{x}$ and propagates along $\hat{y}$.
d. The wave is polarized along $\hat{z}$ and propagates along $\hat{y}$.

1.6.4 When the electric permittivity of a crystal is measured at $\lambda = 1$ μm with respect to an arbitrary Cartesian coordinate system defined by $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$, it is found to be given by the following tensor:

$$\epsilon = \epsilon_0 \begin{bmatrix} 4.786 & 0 & 0.168 \\ 0 & 5.01 & 0 \\ 0.168 & 0 & 4.884 \end{bmatrix}.$$

a. Find the principal dielectric axes $\hat{x}$, $\hat{y}$, and $\hat{z}$ of the crystal and their corresponding principal indices of refraction.
b. Write down the equation that describes the index ellipsoid of the crystal in the original coordinate system. What is the equation for the index ellipsoid in the coordinate system defined by the principal axes?
c. Is the crystal uniaxial or biaxial? Find its optical axis if it is uniaxial or its optical axes if biaxial.
d. How do you arrange an optical wave to propagate in such a crystal so that the polarization of the wave remains unchanged throughout the entire path if the wave is linearly polarized? How about if the wave is circularly polarized?
e. Make a quarter-wave plate for the optical wave at $\lambda = 1$ μm. What is the thickness of the plate?

1.6.5 Under what condition can the polarization of an optical wave propagating in a birefringent crystal remain unchanged for any initial state of polarization and any distance of propagation?

1.6.6 Show that a linearly polarized wave can be converted into a circularly polarized wave by passing it through a quarter-wave plate, and vice versa. In converting a circularly polarized wave into a linearly polarized wave, how do you control

the direction of the linear polarization at the output? Design the arrangement in conducting such an experiment properly in terms of the orientation of the relevant axes and the direction of polarization.

1.6.7 How far must a linearly polarized wave at $\lambda = 1$ μm travel through a crystal that has $n_x = 1.55$ and $n_y = 1.52$ before its polarization is changed into each of the following states. In answering these questions, explain by showing the arrangements with sketches.

a. It is made circularly polarized.
b. It remains linearly polarized but with its polarization rotated by 90°.
c. It remains linearly polarized but with its polarization rotated by 60°.

1.6.8 Quartz is a positive uniaxial crystal, which has n_o=1.544 23 and n_e=1.553 32 at $\lambda = 600$ nm. A quartz plate is cut in such a way that its optical axis is parallel to the surfaces of the plate. A linearly polarized optical beam at 600 nm is sent to pass through such a quartz plate.

a. What is the thickness of a piece of quartz needed to change a linearly polarized beam into a circularly polarized beam at 600 nm wavelength? How should the quartz plate be arranged with respect to the polarization direction of the linearly polarized beam in order for this to happen?
b. What should the thickness of the quartz plate be to enable rotation of the linear polarization of the beam by 50°? How do you arrange the polarization direction with respect to the crystal axes in this case?
c. If instead we want to make sure that the linearly polarized beam stays linearly polarized in the same direction upon passing through the quartz plate irrespective of the polarization direction with respect to the optical axis of the quartz plate, what should the thickness of the plate be?

1.6.9 At what wavelength does a quarter-wave plate for $\lambda = 1$ μm function as a half-wave plate if the dispersion in the refractive indices of the plate is neglected? At what wavelength does light traveling through the plate always return to its input polarization state?

1.6.10 Quartz is a positive uniaxial crystal, which has $n_o = 1.544\,23$ and $n_e = 1.553\,32$ at $\lambda = 600$ nm.

a. Design a quartz waveplate to be used for rotating the polarization direction of a linearly polarized beam at 600 nm wavelength by 60°. Give the thickness of the plate and the arrangement of your setup.
b. If dispersion of the quartz plate can be neglected, at what other wavelengths can this plate be used as a polarization rotator for linearly polarized light?
c. Again, if dispersion can be neglected, at what optical wavelengths can this plate be used to convert a linearly polarized beam into a circularly polarized one?

d. Find the thickness of a plate that has the same function as the one found in (a) if it has to be thicker than 1 mm but thinner than 1.5 mm.

1.6.11 Rutile (TiO_2) is a uniaxial crystal. Its ordinary and extraordinary indices of refraction as a function of wavelength are given by

$$n_o^2 = 5.913 + \frac{0.2441}{\lambda^2 - 0.083}, \tag{1.181}$$

$$n_e^2 = 7.197 + \frac{0.3322}{\lambda^2 - 0.0843}, \tag{1.182}$$

where λ is in micrometers. A rutile plate of thickness l is cut in such a way that its surface normal is perpendicular to its optical axis.

a. If the plate is to be used as a first-order half-wave plate at an optical wavelength of 1 μm, what should its thickness l be? How do you arrange the plate with respect to the polarization of the incident beam if the polarization of a linearly polarized input beam is to be rotated 60° after passing through the plate?

b. With the thickness of the plate obtained in (a), find another wavelength at which the plate also functions as a half-wave plate. Find a wavelength at which it functions as a quarter-wave plate.

1.6.12 Consider wave propagation in a uniaxial crystal whose optical axis is $\hat{z}$.

a. By using the relationships among $\hat{k}$, $\hat{e}_o$, and $\hat{e}_e$ given in (1.121), verify that the unit vectors $\hat{e}_o$ and $\hat{e}_e$ are given by the expressions in (1.123) and (1.124), respectively.

b. By examining the index ellipsoid, show that $n_e(\theta)$ for the extraordinary wave is given by (1.125).

1.6.13 Explain why (1.118) is written in **E** whereas (1.126) is written in **D**. How would **D** be expressed for the wave that is described by (1.118)? Does it have the same form as (1.118)? Why? How would **E** be expressed for the wave that is described by (1.126)? Does it have the same form as (1.126)? Why?

1.6.14 Show that the walk-off angle as defined in Fig. 1.13(*a*) is given by (1.131). Given n_e and n_o for a uniaxial crystal, find the angle θ for the propagation direction $\hat{k}$ that results in the largest walk-off for an extraordinary wave.

1.6.15 An extraordinary optical wave propagates in a uniaxial crystal with its wavevector **k** making an angle θ with respect to the optical axis, $\hat{z}$, of the crystal. In the case when $\theta \neq 90°$, the Poynting vector, **S**, of the wave is not parallel to **k**. The angle α between **S** and **k** is the same as that between **E** and **D**.

a. Show that the vector **S** lies between **k** and the optical axis if the crystal is positive uniaxial, while **k** lies between **S** and $\hat{z}$ if it is negative uniaxial. What is the relationship among **E**, **D**, and $\hat{z}$ in either case?

b. Show that the walk-off angle given by (1.131) can also be expressed as

$$\alpha = \tan^{-1}\left[\frac{n_e^2(\theta)}{2}\left(\frac{1}{n_e^2} - \frac{1}{n_o^2}\right)\sin 2\theta\right], \tag{1.183}$$

where $n_e(\theta)$ is that given by (1.125).

c. Show that the maximum walk-off between **S** and **k** occurs at

$$\theta = \tan^{-1}\frac{n_e}{n_o} \tag{1.184}$$

for

$$\alpha = \tan^{-1}\frac{n_o}{n_e} - \tan^{-1}\frac{n_e}{n_o}. \tag{1.185}$$

1.6.16 Rutile (TiO_2) is a uniaxial crystal. Its ordinary and extraordinary indices of refraction as a function of wavelength are given by (1.181) and (1.182), respectively. A rutile plate of thickness l is cut in such a way that its surface normal is at an angle $\theta = 30^\circ$ with respect to its optical axis. If this plate is used as a polarizing beam splitter for normally incident light at $\lambda = 0.8$ μm, what is the separation between the two orthogonally polarized beams leaving the plate? If the spot size of a collimated incident beam is 100 μm in diameter, what is the minimum value of l for the two orthogonally polarized beams at the exit to be completely separated without spatial overlap?

1.7.1 The intensity profile of a fundamental Gaussian beam, whose field profile is given by (1.138) with $m = n = 0$, at any spatial location is a function of the transverse radial distance, $r = (x^2 + y^2)^{1/2}$, from the beam center and the longitudinal distance z from the beam waist.

a. Show that the intensity profile can be expressed as the following Gaussian function:

$$I(r, z) = I_0(z)\exp\left[-\frac{2(x^2 + y^2)}{w^2(z)}\right] = I_0(z)\exp\left[-\frac{2r^2}{w^2(z)}\right], \tag{1.186}$$

where $I_0(z)$ is the peak intensity of the beam at the longitudinal location z.

b. Express the power, P, of this Gaussian beam as a function of its peak intensity $I_0(z)$ and its spot size $w(z)$ at any location z.

c. Find the variation of the peak intensity $I_0(z)$ along the longitudinal axis of the beam by expressing it as a function of peak intensity I_0 at the beam waist and distance z from the beam waist.

1.7.2 A fundamental Gaussian laser beam of power $P = 1$ W at a wavelength of $\lambda = 532$ nm is focused to a small spot radius of $w_0 = 10$ μm at its beam waist. What is the peak intensity I_0 at the beam waist? What is the divergence angle

of the beam? What are its spot size and peak intensity at a distance of 1 m from the beam waist? If the spot size is reduced by half to $w_0 = 5$ μm at the beam waist, what are the changes of the peak intensities at the beam waist and at 1 m from the waist?

1.7.3 A circular aperture of radius a is placed in the path of a fundamental Gaussian beam with the center of the aperture located at the center of the beam. The Gaussian beam has a spot size w at the location of the aperture.

a. Find and plot the fraction of beam power transmitted through the aperture as a function of a and w.
b. What percentage of power is transmitted if the aperture has a radius equal to the beam spot size, $a = w$?
c. What is the minimum aperture diameter for the aperture to transmit at least 99% of the beam power?

1.7.4 A laser retroreflector was first placed on the lunar surface by the astronauts of the *Apollo 11* lunar landing mission in 1969. Similar retroreflectors were placed on different parts of the lunar surface by astronauts in later missions, including *Apollos 14* and *15*. These retroreflectors have since been used for precision lunar laser ranging to measure the distance between Earth and the Moon using nanosecond and picosecond laser pulses down to a precision of the order of 1 cm. The *Apollo 11* retroreflector consists of an array of 100 silica corner cubes in a 46 cm × 46 cm panel. Each corner cube has a diameter of 3.8 cm. The function of a corner cube is to reflect the light intercepted by it right back to the original direction from which the light comes without the need for critical alignment. The distance between the centers of Earth and the Moon is about 385 000 km, but the direct distance between their surfaces is shorter and is about the distance for light to travel in 1.25 s. In this problem, we consider a lunar ranging experiment using a telescope of 1.5 m diameter to collimate laser pulses of 350 ps duration at a wavelength of 532 nm from the second harmonic of a Nd : YAG laser. We assume that the laser beam has a fundamental Gaussian profile of waist spot size $w_0 = 0.5$ m at the aperture of the 1.5-m-diameter telescope. We also assume that each corner cube in the retroreflector reflects about 80% of the laser light it intercepts but adds a divergence of 14 μrad to the reflected beam due to diffraction. In answering the following questions, we first ignore the scattering, absorption, diffraction, and dispersion caused by the atmosphere.

a. What is the divergence angle of the out-going beam? What is the spot size of the beam on the Moon's surface?
b. If the laser beam is incident on the retroreflector with the beam center located at the center of the panel, what fraction of the laser energy is intercepted and reflected back by the retroreflector?

c. What is the spot size of the reflected beam on Earth? What fraction of the beam reflected by the retroreflector back to Earth is intercepted by the 1.5-m receiving aperture of the telescope?
d. What fraction of the energy in each laser pulse is finally received after making the round trip to the Moon and back? If we hope to detect at least one photon in each pulse, what is the minimum energy required of the original out-going pulse?
e. In reality, the effects of the atmosphere are significant and certainly cannot be ignored unless the entire station is located in space. In each pass, the atmosphere adds a divergence of about 18 μrad to the beam mainly due to dispersion and transmits only about 2% due to scattering and absorption. Answer questions (a)–(d) with the atmospheric effects accounted for.

1.7.5 The effect of sending a Gassian beam through a thin lens of focal length f can be described by the following relation:

$$\frac{1}{q'} = \frac{1}{q} - \frac{1}{f}, \tag{1.187}$$

where q and q' are the complex radii of curvature of the Gaussian beam immediately before and after the thin lens. The value of f can be either positive or negative for a positive or negative lense, respectively. A Gaussian beam of waist radius w_0 located at $z = 0$ is sent through a thin lens of focal length f located at $z = z_0$.

a. Show that the waist radius for the beam after passing through the lens is

$$w_0' = \frac{|f|}{[(z_0 - f)^2 + z_R^2]^{1/2}} w_0, \tag{1.188}$$

where z_R is the Rayleigh range of the incoming beam.
b. Show that the waist of the beam passing through the lens is located at

$$z = \frac{z_0^2(z_0 - f) + z_R^2(z_0 + f)}{(z_0 - f)^2 + z_R^2}. \tag{1.189}$$

c. How can the beam be best collimated? What are the waist radius and Rayleigh range of this optimally collimated beam?
d. If the lens is placed at the waist location of the incoming beam, what is the waist radius of the outgoing beam? Where is the waist located? What is the effect of the lens on the divergence of the beam?

1.8.1 Under what condition is an optical wave that is reflected from a dielectric surface completely polarized no matter whether the incident wave is polarized or not? What is its state of polarization?

1.8.2 A beam of circularly polarized light is incident from the air on the surface of an isotropic lossless dielectric material. The index of refraction of the dielectric material is unknown. However, it is found experimentally, by varying the angle of incidence, that the reflected light is linearly polarized when the angle of incidence is 60°. What is the index of refraction of the dielectric material? Explain what happens.

1.8.3 A reflection-type polarizer can be made simply with a glass plate. If the glass plate available has an index of refraction $n = 1.5$ at the wavelength of interest, what should the incident angle of the light be in order for the plate to function as a polarizer? Illustrate how this device should be used if the incident light is arbitrarily polarized.

1.8.4 During a sunny day on the equator when the sun rises at 6 a.m. and sets at 6 p.m., at what times is the sunlight reflected from the ocean surface most polarized?

1.8.5 When sunlight reflected from the surface of a body of water is viewed through linearly polarizing glass, the apparent glare from the water is reduced.

a. Upon which concept discussed in this chapter is this glare reduction based?

b. Suppose you have a beachfront house, and you want to use polarizing glass to reduce the glare from the sunlight reflected from the ocean. How should you orient the polarizing glass? (Should the glass block horizontally or vertically polarized light?)

c. For what angle of reflected sunlight will your polarizing glass be most effective? (Assume that the index of refraction of water is 1.33.)

1.8.6 The index of refraction of ordinary glass is $n = 1.5$.

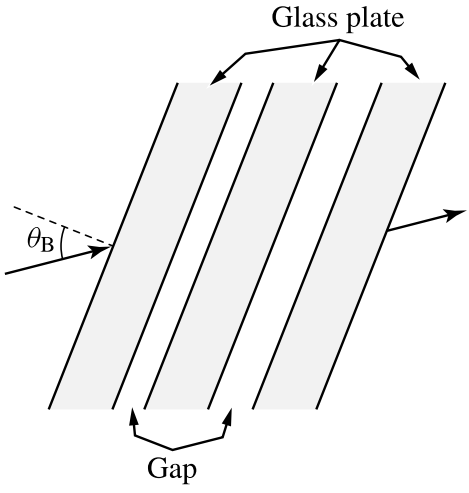


Figure 1.23 Stack of parallel flat glass plates.

a. Find the Brewster angles for the incidence of light from air to glass and from glass to air, respectively. What is the angle for total internal reflection?

b. For a stack of parallel flat glass plates separated by air gaps as shown in Fig. 1.23, show that if TM-polarized light is incident on the surface of the first plate at the Brewster angle, it is transmitted through the whole stack without reflection at any interface. Sketch the path of light through the stack. What are the effects of the thickness of the plates and that of the air gaps?

c. What happens if one air gap is filled with water whose index of refraction is 1.33? Illustrate by sketching the path of light.

1.8.7 The indices of refraction for water and diamond are 1.33 and 2.42, respectively.

a. For a piece of diamond exposed to the air, what are the critical angle for internal reflection, the Brewster angle for external reflection, and the reflectivities for TE and TM waves at an incident angle of 45°?

b. Answer the same question for a piece of diamond that is immersed in water.

1.8.8 A 90° symmetric prism with antireflection coating at the input surface as shown in Fig 1.24 can be used as a retroreflector. This kind of prism can losslessly reflect light with an adjustable lateral displacement between the paths of the incident and reflected beams.

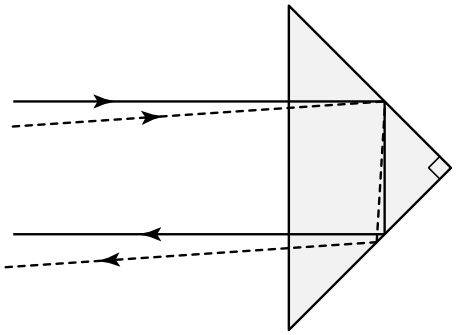


Figure 1.24 Prism retroreflector.

a. Show that the path of the reflected beam is parallel to that of the input beam for both normal and oblique incidence, thus requiring no critical alignment.

b. However, if the angle of incidence is too large, the reflected beam will suffer losses. What is the condition for a retroreflecting prism to have an angular tolerance of $\pm 5^\circ$ with respect to normal without substantial loss?

1.8.9 At the optical wavelength of 500 nm, GaAs is measured to have a reflectivity of 40% at normal incidence and an absorption coefficient of $\alpha = 10^7\ \mathrm{m}^{-1}$.

a. What is the complex refractive index of GaAs at 500 nm? What is the complex susceptibility?

b. Plot the reflectivity of GaAs at 500 nm as a function of incident angle for both TE and TM polarizations. What is the lowest reflectivity for the TM polarization? At what incident angle does it occur?

1.9.1 Explain how the primary rainbow is formed and describe the sequence of the rainbow colors from top to bottom. Answer the same questions for the secondary rainbow and explain the differences between the primary and secondary rainbows. Explain also why a rainbow has the shape of an arc. Find the arc angles for the primary and secondary rainbows.

1.9.2 The $LiNbO_3$ crystal is negative uniaxial. Its indices of refraction for the ordinary and extraordinary waves at room temperature as a function of optical wavelength are given by the following Sellmeier equations:

$$n_o^2 = 4.9130 + \frac{0.1188}{\lambda^2 - 0.045\,97} - 0.0278\lambda^2, \tag{1.190}$$

$$n_e^2 = 4.5798 + \frac{0.0994}{\lambda^2 - 0.042\,35} - 0.0224\lambda^2, \tag{1.191}$$

where λ is in micrometers. For both ordinary and extraordinary waves at an optical wavelength of 1.3 μm, find (a) the phase velocities, (b) the group velocities, and (c) the group-velocity dispersion parameters.

1.9.3 The BBO crystal is negative uniaxial. Its indices of refraction for the ordinary and extraordinary waves at room temperature as a function of optical wavelength are given by the following Sellmeier equations:

$$n_o^2 = 2.7359 + \frac{0.018\,78}{\lambda^2 - 0.018\,22} - 0.013\,54\lambda^2, \tag{1.192}$$

$$n_e^2 = 2.3753 + \frac{0.012\,24}{\lambda^2 - 0.016\,67} - 0.015\,16\lambda^2, \tag{1.193}$$

where λ is in micrometers. For both ordinary and extraordinary waves in the optical wavelength range between 0.5 and 2.0 μm, plot (a) phase velocity, (b) group velocity, and (c) group-velocity dispersion, as a function of wavelength.

1.10.1 Find the exact $\chi(\omega)$ corresponding to $\chi(t)$ given in (1.174) without making the rotating-wave approximation used in (1.175). Show that the real and imaginary parts of this exact $\chi(\omega)$ are even and odd functions of ω, respectively. Compare them with their respective approximate expressions in (1.176) to justify the applicability of the latter. Show that the exact expression for $\chi(\omega)$ satisfies the reality condition, as expected.

1.10.2 A material has two closely spaced resonance frequencies at ω_{01} and ω_{02} with respective response constants χ_{b1} and χ_{b2} and relaxation constants γ_1 and γ_2. The condition $0 \ll \omega_{02} - \omega_{01} \ll \omega_{01}$ is always valid in this problem.

a. Consider the case when $\chi_{b1} = \chi_{b2}$ and $\gamma_1 = \gamma_2 = \omega_{02} - \omega_{01}$. Sketch the real and imaginary parts of $\chi(\omega)$ as a function of ω near the two closely spaced

resonance frequencies. Also indicate the regions of normal and anomalous dispersion.

b. What changes to the dispersion of a laser material do you expect when its resonance at ω_{01} is pumped to population inversion but not that at ω_{02}, meaning that χ_{b1} changes sign but χ_{b2} does not? Sketch the real and imaginary parts of $\chi(\omega)$ as a function of ω near the two closely spaced resonance frequencies in this situation.

c. Sketch the real and imaginary parts of $\chi(\omega)$ as a function of ω near the two closely spaced resonance frequencies in the situation when population inversion occurs at both resonances so that both χ_{b1} and χ_{b2} change sign. Indicate the regions of normal and anomalous dispersion.

d. Answer questions (a)–(c) for the case when $\chi_{b1} = 3\chi_{b2}$ but $\gamma_1 = \gamma_2/3 = \omega_{02} - \omega_{01}$.

1.11.1 What is the separation in energy between the two energy levels that are responsible for emission at $\lambda = 1.064$ μm of a Nd : YAG laser?

1.11.2 A ruby is basically crystalline Al_2O_3 doped with Cr^{3+} impurities. Its red color is caused by the fact that the Cr^{3+} ions emit light at 694.3 nm when making the transition from an excited state to the ground state. What is the energy level of this excited state?

1.11.3 Silicon has a bandgap of 1.12 eV at room temperature. What wavelengths in the optical spectrum are transmitted through a pure silicon wafer?

1.11.4 GaAs has an energy bandgap of 1.424 eV at room temperature and absorbs any photon that has an energy higher than this value. For what optical wavelengths is GaAs transparent?

1.11.5 Consider monochromatic light illuminating a photographic film. The incident photons will be recorded if they have enough energy to dissociate the AgBr molecules in the film. The minimum energy required to do this is about 0.6 eV. Find the cutoff wavelength longer than which the incident light will not be recorded. In what spectral region does this wavelength fall?

1.11.6 A photon of 10.6 μm wavelength is combined with a photon of 1.06 μm wavelength to create a photon that combines the energies of both. What is the wavelength of the resultant photon?

SELECT BIBLIOGRAPHY

Born, M. and Wolf, E., *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th edn. Cambridge: Cambridge University Press, 1999.

Fowler, G. R., *Introduction to Modern Optics*, 2nd edn. New York: Dover, 1975.

Haus, H. A., *Waves and Fields in Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

Iizuka, K., *Elements of Photonics in Free Space and Special Media*, Vol. I. New York: Wiley, 2002.

Jackson, J. D., *Classical Electrodynamics*, 3rd edn. New York: Wiley, 1999.

Landau, L. D. and Lifshitz, E. M., *Electrodynamics of Continuous Media*. Oxford: Pergamon, 1960.

Nye, J. F., *Physical Properties of Crystals*. London: Oxford University Press, 1957.

Parker, S. P., *Optical Source Book*. New York: McGraw-Hill, 1987.

Saleh, B. E. A. and Teich, M. C., *Fundamentals of Photonics*. New York: Wiley, 1991.

Sirotin, Yu. I. and Shaskolskaya, M. P., *Fundamentals of Crystal Physics*. Moscow: Mir Publishers, 1982.

Yariv, A. and Yeh, P., *Optical Waves in Crystals: Propagation and Control of Laser Radiation*. New York: Wiley, 1984.